

Result No.	δ			ID	Description
	Score	Match	Length		
1	144.8	7.6	93695	7	AL021749 Arabidops

```
/db_xref="GI:2842478"
```

/db_xref="GI:2842478"

/gene="F2009.30"

```
/number=4
complement(7537..7647)
/gene="F2009.30"
/number=5
complement(7648..7762)
/gene="F2009.30"
/number=5
complement(7763..7915)
/gene="F2009.30"
/number=6
complement(7916..9900)
/gene="F2009.30"
/number=6
complement(9901..9975)
/gene="F2009.30"
/number=7
complement(9976..10604)
/gene="F2009.30"
/number=7
10056..10127
/gene="F2009.30"
/note="Lys-tRNA(TTT)"
complement(10605..10996)
/gene="F2009.30"
/number=8
complement(10997..11073)
/gene="F2009.30"
/number=8
complement(11074..11254)
/gene="F2009.30"
/number=9
13176..17409

Query Match          7.6%; Score 144.8; DB 7; Length 93695;
Best Local Similarity 65.5%; Pred. No. 1.1e-09;
Matches 228; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

Qy 724 agcgctggcagagcacctctcgagggcggtgcgtacatgctatcgccgagcagctc 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33691 AGGGGGTGGCAGATTAGTGAACCGAGACTTGACGACAGGTAAGGCTTAAGCTA 33750

Qy 784 ttctccaccgcggcgagccaccagcgctcaggttcataatcccggtgctggccagacc 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33751 TTCAATACTGCGGATGCAACCAAGGATAGAAGTAGTTCCTCAGTCAGTGGCTCGACCA 33810

Qy 844 gccggcgcaacattctctcccagggccaggtaccacaaactacagggcgccgcgcg 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33811 AA---CGTAAACATCCTTCTCCCAAGGCTAGTTACCTCCTACGAGGCTCGTCCGTC 33867

Qy 904 ttcaacaggtgaggtcgccgattctcgacctcattcccgacaaagggtggagatcgac 963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33868 TACAGTGACTCGAGTCCGCAAGTTCGATCTCCTTCCAGAGAAAGATGGAGATTGAT 33927

Qy 964 atcgactcgctggaaatccatcgccgacagaacacccacccgcatggttcataaacccc 1023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33928 CTCCAGGCGATCGAAGCCATGGCAGATGAGAACACAGTTGCAATGGTTATCATAAACCT 33987

Qy 1024 aacaaccgctggcagcggttactcctacgaccatctgtccaaagtc 1071
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33988 AACAAACCCCTGTGGAATGTTTACTCTTACGATCATCTTAAAGAGGTC 34035

RESULT 2
ATAC006585
LOCUS ATAC006585 103495 bp DNA PLN 04-APR-1999
DEFINITION Arabidopsis thaliana chromosome II BAC F27C12 genomic sequence,
complete sequence.
ACCESSION AC006585
NID 94559344
VERSION AC006585.6 GI:4559344
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis
1 (bases 1 to 103495)
Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
Barnstead,M.E., Mason,T.M., Bowman,C.L., Renning,C.M., Benito,M.,
Carreira,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
Fraser,C.M. and Venter,J.C.
Arabidopsis thaliana chromosome II BAC F27C12 genomic sequence
Unpublished
2 (bases 1 to 103495)
Lin,X. and Kaul,S.
Direct Submission
Submitted (19-FEB-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 103495)
Lin,X.
Direct Submission
Submitted (04-APR-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Apr 4, 1999 this sequence version replaced gi:4522000.
Address all correspondence to:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org
BAC clone F27C12 is from Arabidopsis chromosome II and is near the
molecular marker mi238.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tadb/at/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.
FEATURES
source
1..103495
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="mi238"
/clone="F27C12"
100..136
/repeat_region
/repeat_family="AT_rich"
206..243
/repeat_family="AT_rich"
complement(245..344)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
262..349
/note="exon predicted by xgrail, quality excellent"
complement(353..418)
/repeat_family="AT_rich"
join(1069..1197,1624..1742,1826..1889,2061..2169,
mrna
```

2257. .2410,2506. .2755,2850. .2926,3010. .3062,3149. .3203,3287. .3608)
/gene="F27C12.1"
1069. .3608
/gene="F27C12.1"
join(1626. .1742,1826. .1889,2061. .2169,2257. .2410,2506. .2755,2850. .2926,3010. .3062,3149. .3203,3287. .3475)
/gene="F27C12.1"
/codon_start=1
/product="protein phosphatase 2C"
/protein_id="AAD23006.1"
/db_xref="PIR:G4559345"
/db_xref="GI:4559345"
/translation="MGTVLSSPKTEKLSGDSNKLRFGLSMQGWRAITMEDAAHAAIL
DLDUITSFFVYDGGVKAQVAKFAKYLHQVYISNEAYKTGDVTSLRARFPRMDMM
QGQWRRELAVLGDKNFGSMGIEGFIWSPSGDITNPDSPLEDPGPHSDFTGPTSG
CTACVALLKDKLFFVANKGDSKCVISQKQYANLSKDKHPDLEVEKKERILKAGGIHA
GRINGSLNLTGAIGDMERKQNFPEKQMTADPINTILDCDDKEDILFVACDGIWD
CMSQELVDFTHEQLKSTKTSKTCGEKVVYDCLAPDTGEGCDNMTILVQFKKPNP
SETPEDSKPESEDEPSSSS"
3939. .3988
/note="exon predicted by xgrail, quality
excellent_shadowexon"
complement(join(<4062. .4414,4489. .>4666))
/gene="F27C12.2"
complement(<4062. .>4666)
/gene="F27C12.2"
complement(join(4062. .4414,4489. .4666))
/gene="F27C12.2"
/codon_start=1
/product="early nodulin 16"
/protein_id="AAD23007.1"
/db_xref="PIR:G4559346"
/db_xref="GI:4559346"
/translation="MASSSLHVAIFSLFLFLSLAANEVTVGKSGDKWKIPSSSYSF
TEWAQKRAKVGDFIVFRYEGSKDVLTEKYNSCTNPPLANTDGETKVKLDLRS
GPFYFISAGANGCHCGKQLSLVLSPRHSVISAPSPVEFDGPAAPAPISGVSRLG
GCYVVLGIVLIGLCANF"
complement(5439. .5548)
/note="exon predicted by xgrail, quality good"
5777. .5857
/rpt_family="Af_rich"
complement(5974. .6004)
/note="exon predicted by xgrail, quality excellent"
complement(6006. .6052)
/rpt_family="Af_rich"
complement(6248. .6315)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(6332. .6545)
/note="exon predicted by xgrail, quality good"
6795. .6873
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(7159
7064. .7159
/note="exon predicted by xgrail, quality marginal"
join(<7609. .8283,8382. .8603,8902. .>10233)
/gene="F27C12.3"
<7609. .>10233
/gene="F27C12.3"
join(7609. .8283,8382. .8603,8902. .10233)
/gene="F27C12.3"
/codon_start=1
/product="putative extensin"
/protein_id="AAD23008.1"
/db_xref="PIR:G4559347"
/db_xref="GI:4559347"
/translation="FMLEDYRVYVSRIMSOLREOPGASFWFNFRDGRSDRSRMESV
LTYEDMTIMDYPRHVEGCPILTMETVHFLKSAEWLLLSOONILLSCELGGPPTLA
FMVLLLTIRKQFSGEHTLLEMIYKQAPRELLQLMSPLNLPQSURLFQYISRRNVS
QWPLDQALITLDCVNLRLIPDFDEGGCRPIFRYIGODPFWFASTKYSFMPKRSK
AVRYKQADCLVLDINDCHILGDLVLCITLGDLSLEEMMFVRYFTAFPLRSNLIT
LNRGDILVNLNITDRFPGDAEYVIFSEMGAQKLSAVDLPHMEKDVLPWEAFKVO

[illegible]

euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS

1 (bases 1 to 61384)
Lin, X., Kaul, S., Shea, T.P., Fujii, C.Y., Shen, M., VanAken, S.E., Barnstead, M.E., Mason, T.M., Bowman, C.L., Ronning, C.M., Benito, M., Carrara, A.J., Creasy, T.H., Buell, C.R., Town, C.D., Nierman, W.C., Fraser, C.M. and Venter, J.C.

TITLE
JOURNAL

Arabidopsis thaliana chromosome II BAC F23N11 genomic sequence

REFERENCE
AUTHORS

2 (bases 1 to 61384)

TITLE
JOURNAL

Lin, X. and Kaul, S.
Direct Submission
Submitted (09-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org

REFERENCE
AUTHORS

3 (bases 1 to 61384)

TITLE
JOURNAL

Lin, X.
Direct Submission
Submitted (25-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

COMMENT

On Mar 25, 1999 this sequence version replaced gi:4389529.
Address all correspondence to:

Xiaoying Lin

The Institute for Genomic Research

9712 Medical Center Dr.

Rockville, MD 20850, USA

e-mail: xlin@tigr.org

BAC clone F23N11 is from Arabidopsis chromosome II and is near the

molecular marker mil148.

The orientation of the sequence is from SP6 to T7 end of the BAC

clone.

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

FEATURES
SOURCE

1. .61384
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="mil148"
/clone="F23N11"
1
/note="16680 bases before this point were not included in the submitted sequence, but they appear in the overlapping BAC."

misc_feature

1
/note="exon predicted by xgrail, quality marginal_shadowexon"

misc_feature

1. .13161
/note="overlap with BAC clone F5H14 (AC006234.116508..129667)."

repeat_region

complement(59..103)

repeat_region

384..409
/rpt_family="AT_rich"

repeat_region

697..721
/rpt_family="AT_rich"

repeat_region

1012..1035
/rpt_family="AT_rich"

misc_feature

1433..1572
/note="exon predicted by xgrail, quality excellent"

misc_feature

complement(1528..1620)
/note="exon predicted by xgrail, quality marginal_shadowexon"

mRNA

join(<1719..1934,2049..2277,2589..>3028)

gene

/gene="F23N11.1"

<1719..>3028

/gene="F23N11.1"

/note="F5H14.36"

join(1719..1934,2049..2277,2589..3028)

/gene="F23N11.1"

/note="unknown protein"

/codon_start=1

/protein_id="AAD21702.1"

/db_xref="PIR:G4512647"

/db_xref="GI:4512647"

/translation="MDRQNSDDIMRFLDGMASDDVLFGLDEGNQSPEDFSVNLNAG
EDDDEDDNNNSDNKAFQEHQELQGLTYRTSSLETKIQATKQALKQVSKGLY
CVCRRPVDGGSCSLRGSEISHLRDVAGYDCVSKSKWRCQDIPAGEHEFIEIVDRS
GSKSEMRVVIELSFRAEFAKGESEYKRLISRPVYGVKTERLSRLKILKILGACG
KCLRDKKHMAPWRKHRYMQAKLGTCDRSSLSLEASVSEAMEPENWVPAKPRVSMNLN
YDGLLGGFSAGPATVAVV"

complement(2296..2324)

/rpt_family="AT_rich"

complement(2565..2585)

/rpt_family="AT_rich"

complement(3117..3201)

/note="exon predicted by xgrail, quality marginal_shadowexon"

repeat_region

3327..3389
/rpt_family="POLY_A"

misc_feature

3837..3973

/note="exon predicted by xgrail, quality marginal"

complement(4252..4308)

/rpt_family="(TA)n"

4307..4352

/note="exon predicted by xgrail, quality good"

complement(4870..4893)

/rpt_family="AT_rich"

5283..5408

/rpt_family="POLY_A"

complement(<5419..>5724)

/gene="F23N11.2"

complement(<5419..>5724)

/gene="F23N11.2"

/note="predicted by genscan"

complement(5419..5724)

/gene="F23N11.2"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD21703.1"

/db_xref="PIR:G4512648"

/db_xref="GI:4512648"

/translation="NKLIFAVIISVWLFPPVLVSSRTIKCDOLSGKINGEKEITMNM
RLGLDVSSRRILQASRYISYEALKKLNLDNRGEPDQRPYRRSCDVHSHCYRFTN"

5852..5890

/note="exon predicted by xgrail, quality marginal_shadowexon"

6449..6487

/rpt_family="AT_rich"

6837..6880

/rpt_family="AT_rich"

7043..7110

/rpt_family="POLY_A"

7231..7292

/rpt_family="AT_rich"

7374..7406

/rpt_family="AT_rich"

complement(7723..7789)

/rpt_family="AT_rich"

misc_feature	complement(7930..8000)	
misc_feature	/note="exon predicted by xgrail, quality good_shadowexon" 8174..8302	
misc_feature	/note="exon predicted by xgrail, quality excellent" complement(8465..8586)	
misc_feature	/note="exon predicted by xgrail, quality good_shadowexon" 8544..8629	
misc_feature	/note="exon predicted by xgrail, quality good" 8702..8829	
misc_feature	/note="exon predicted by xgrail, quality good" complement(8934..9057)	
misc_feature	/note="exon predicted by xgrail, quality good_shadowexon" 8936..9051	
misc_feature	/note="exon predicted by xgrail, quality excellent" 9147..9243	
misc_feature	/note="exon predicted by xgrail, quality excellent" 9290..9338	
misc_feature	/note="exon predicted by xgrail, quality marginal" complement(9317..9461)	
misc_feature	/note="exon predicted by xgrail, quality good_shadowexon" 9492..9555	
misc_feature	/note="exon predicted by xgrail, quality marginal" complement(9634..9780)	
misc_feature	/note="exon predicted by xgrail, quality marginal_shadowexon" 9666..9814	
misc_feature	/note="exon predicted by xgrail, quality excellent" complement(9912..9988)	
misc_feature	/note="exon predicted by xgrail, quality marginal_shadowexon" join(<10040..10154,10340..10507,10608..10936,11019..>11237)	
misc_feature	/gene="F23N11.3" <10040..>11237	
misc_feature	/gene="F23N11.3" /note="F23N11.3"	
misc_feature	/note="F5H14.37; predicted by genscan" join(10040..10154,10340..10507,10608..10936,11019..11237)	
misc_feature	/gene="F23N11.3" /note="hypothetical protein" /codon_start=1	
misc_feature	/protein_id="AAD21704.1"	
misc_feature	/db_xref="PID:g4512649"	
misc_feature	/db_xref="GI:4512649"	
misc_feature	/translation="MTLIVYKGAARVSTMIGQAAIMDSYLCILHLTAGILVESLFNA"	

Query Match	7.0%	Score 134.2;	DB 8;	Length 61384;
Best Local Similarity	63.7%;	Pred. No. 2.1e-08;		
Matches 221;	Conservative 0;	Mismatches 123;	Indels 3;	Gaps 1;
QY 724	agggccgtggcagagaccctgtcgcagggcgtgcggtacatgctatcgccgcagacgcctc	783		
Db 35491	AGGCCGTGCTGAATATATGAACCGAGATCTCCGCACAAAGTTAACGCTTGAAGATATT	35550		
QY 784	tctctaccgcgcggcgagaccagcgcatcgaggtcataatcccggtgctggtcccgagacc	843		
Db 35551	TTTCGACCCGTGGATGCAACCAAGGATAGAGATCGTGTGCAATCGTGGCTCGACCA	35610		
QY 844	gcggcgcccaacatctctgctctcccagggccaggtctacccaactcagagggcgcgccgcg	903		
Db 35611	AA---CGCAAAACTTTGCTCCCAAGTCTGGCTTCCTCACTACGACGCTCGTCTGCT	35667		
QY 904	ttaacacggctggaggtccggcatttcgacctcatcccccaaggggtggagatcgac	963		
Db 35668	TACAGTGGTCTCGAGGTTGCGAAGTTTGATCTTCTCCCGAGAAGAATGGGAGATTGAT	35727		
QY 964	atcgactcgtgggaatccatctgcgcgaagaacaccaccgccttggtcatcataaaacccc	1023		
Db 35728	CTTGAAGGTATCGAAGCCATTGACAGCAGAACACTGTGGCTATGGTGTGAATTAAACCC	35787		
QY 1024	aaacaccgcgtgcgcagcgtttactctctaagaccatctgtccaaaggt	1070		
Db 35788	AACAATCCCTGTGGAAATGTCCTACTCTCAGACCACTCTCAAAAAGGT	35834		

```

RESULT 5
RATTATRA
LOCUS      RATTATRA             2004 bp      mRNA           ROD
DEFINITION Rat tyrosine aminotransferase mRNA, complete cds.           15-MAR-1990
ACCESSION M8340 K01265
VERSION    M8340.1 GI:207155
KEYWORDS   tyrosine aminotransferase.
SOURCE     Rat (adult Wistar) hepatocyte, cDNA to mRNA, clones pCTAT[1-4].
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 2004)
            Hargrove,J.L., Scoble,H.A., Mathews,W.R., Baumstark,B.R. and
            Blemann,K.
TITLE      The structure of tyrosine aminotransferase: Evidence for domains
            involved in catalysis and enzyme turnover
JOURNAL    J. Biol. Chem. 264, 45-53 (1989)
MEDLINE    89079691
COMMENT    Draft entry and computer-readable sequence for [1] kindly submitted
            by J.L.Hargrove, 07-NOV-88.
FEATURES   Location/Qualifiers
            1..2004
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            98..1462
            /note="tyrosine aminotransferase (EC 2.6.1.5)"
            /codon_start=1
            /protein_id="AAA42203.1"
            /db_xref="PID:g207156"
            /db_xref="GI:207156"
            /translation="MDSYVITQTDSDLSVLVDVHVNIGRNSVQCKRKRARWDVR
            PDSMSTNPIRAKTDNKNKVPNPKTVISLIGDPTVEGNLPTDPEVTOAMKDAL
            SKNYGVAISIGLSREVASVYHCHAEPLAKDVLISGGSQAIECLAVLNPQG
            NLLIPRSLYTLAEIKVLEKYNLLPEKSEIDLKQLESLIDKEKLECLVNNPS
            NPGSVFSKRRHKQILVAERQCPILADEIYGDMDVFSDCKVEPLANLSTNPVILSCG
            POBFYHDTLSFKSNADLCYALADLPQLQVPYPSGAMYLVMVGIEMHFEPENDVEF
            TERLIAEQAVHCLPATCFEVPNFVITVPEVMMLEACSRIQEFCEQHYHCDGSGSE
            ECDK"
BASE COUNT      491 a      521 c      517 g      475 t
ORIGIN

Query Match      6.9%; Score 131; DB 12; Length 2004;
Best Local Similarity 49.1%; Pred. No. 1e-07;
Matches 493; Conservative 0; Mismatches 500; Indels 12; Gaps 5;

QY 549 cagcgagcgtgcagagagaagggcgccgctgctgcgcctgccccacgggaccc 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 CAACATGAAGTGCAGGCCCAATCCGACAAAGACCGTGATTCTCTGTCAATTGGGACCC 339

QY 609 gtccgtgtcccgcccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgt 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 TACTGTGTTGGNACCTGCTACAGACCCTGAAGTACCACAGCCATGAAGATGCCCT 399

QY 669 gcgcacggccagttcaactgtctaccccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 GGACTCGGGGAAGTACAATGGCTATGCCCGCTCATCGGCTACCTATCCAGTCGGGAGGA 459

QY 729 cgtggcagagcacctgtcgcagggcgtgcgtacatgctatcgcgcgcgcgcgcgcgcgcgcgt 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 GGTGCGTTCATTAC---TACCACATGTCAATGAGGCTCCTCTGGAAGCATGAAGATGCTATTCT 516

QY 789 caccgcgcggcgagaccagcgatcgagggtcataaatacccggtgctgcgcgcgcgcgcgcgcgcgcgc 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 GACAAAGCGGTGCAGTCAGGCCCATGAGCTATGCTAGTGTGTTGGCC---AATCCTGG 573

QY 849 cgccaacattctctccccaggccaggtctacccaaactacagggcgcgcgccgcttcaa 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 ACAAAACATCCTCATTCCAAGGCCCGGGTTTCCCTCTATAGGACTTTGGCTGAGTCTAT 633

```


Qy	1151	tcacccaatgggagtggt----tgggcacatcacccctgtgtctgtcgaatgaggtctctctgt	1207
Db	33865	ACATCCCCCTGGCGTCTCCCGGGATGCGGGAGCGACAGGTACCATCGGTGGCGCG	33924
Qy	1208	ccaagtcgatgatgtagtcgtggatggcggttgatgaggtagcgg	1252
Db	33925	CCAGAGAGCTTCTCTACACCGCTGGGAAGCTCGCTGGGTACCG	33969

RESULT	7
LOCUS	HSTATR 2754 bp mRNA PRI
DEFINITION	Human mRNA for tyrosine aminotransferase (TAT) (EC 2.6.1.5).
ACCESSION	X52520
NID	g36712
VERSION	X52520.1 GI:36712
KEYWORDS	aminotransferase; transferase; tyrosine aminotransferase.
SOURCE	human

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2754)

REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 2754)
Scherer, G.
Direct Submission
Submitted (08-MAR-1990) Scherer G., Institute of Human Genetics,
Albertstr 11, D 7800 Freiburg

REMARK	JOURNAL
9bases 1-2754)	
2 (bases 1 to 2754)	
REFERENCE	
AUTHORS	
TITLE	
Isolation and characterization of the human tyrosine amino transferase gene	
Nucleic Acids Res. 18 (13), 3853-3861 (1990)	

90326506
MEDLINE
COMMENT
See <X52507>-<X52520>.
Data kindly reviewed (26-JUL-1990) by G. Scherer.

FEATURES	Location/Qualifiers
source	1. 2754
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16q22.1"
	/tissue_type="placenta (genomic) and liver (cdna)"
	/clone_lib="genomic (phage) + cdna"
	/clone="lambda-hrAT1, phcTAT2-16 + phcTAT3a-6"
CDS	97..1461

CDS

97..1461

/note="tyrosine aminotransferase (AA 1-454)"

```

/codon_start=1
/protein_id="CAA36750.1"
/db_xref="PID:g36713"
/db_xref="GI:36713"
/db_xref="SWISS-PROT:P17735"
/translation="MDPYMIOMSSKGNLPSILDVHVNVGRSSYPGKMKGRKRWK
PSDMAKTFNPIRAIVLDMNWKVPNKNTMISLIGDTPVFNGLTDPDPTVQAMKD
SGKYNVAPSGTGFSSREELASYHCPEAPLEAKDVILTSCGQAIDCLCLAVLAN
NILLPRFGSLYKTLAESNGVLEINLNGLPKSEWEIDKQLEILIDEKTLACLVNV
NPGCVSKRHLQKILAAVAARQCPVILADEYIGDMVDFSDKEPIELATLTDVPIIL
GLAKRWLPGWRLGWLILHDIRDFGNEIRDLGVKLSQRILGPGCTIYVGALKSTIL
PFEYVHNLTSLFKSNADICGALAAIPGLRVPVPSGMYLWVGMEHMEHPFENED
TERLVAEGSVHCLPATCFEYPNFIIRVVITVPEVMMLEACSRIOQFCEHYHCAEQ
ECDEK"

```

```

misc_feature      2070..2075      mRNA polyadenylation signal
                  /note="minor mRNA polyadenylation signal"
polyA_site        2090
                  /note="minor mRNA polyadenylation site"
repeat_region     2116..2130
                  /note="direct repeat flanking Alu element"
                  complement(2131..2479)
misc_feature       /note="Alu element"
repeat_region     2480..2494
                  /note="direct repeat flanking Alu element"
misc_feature       2738..2743
                  /note="major mRNA polyadenylation signal"
polyA_site        2754
                  /note="major mRNA polyadenylation site"

```

[illegible]

RESULT	8	
HSTYAT		
LOCUS	2051 bp	13-JUL-1995
DEFINITION	H sapiens mRNA for tyrosine aminotransferase.	
ACCESSION	X56675	
NID	g37501	
VERSION	X56675.1	
KEYWORDS	tyrosine aminotransferase.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	1 (bases 1 to 2051)	
TITLE	Labrie,F.	
JOURNAL	Direct Submission	
	Submitted (26-SEP-1990) Labrie F., CHUL Research Centre, Labo	

Db 50937 TTTATACATCGGGTTGCACGCAAGCATCGATGATAGCATTTGCGATGTTAGCTCGTCC 50996
QY 844 gcygcgcacaacatctgtctcccccagcgagctaccacaaacacagcgagcgccgcgcg 903
Db 50997 A---GGGCTAAATACTTCTTCCAAGGCTGTGTTTCCCAATCTATGAATCTGTCTAAG 51053
QY 904 ttcaacagctgaggtcgcgattctgacccatcccccagcaaggggtggagatcgac 963
Db 51054 TTTAGACACTTGAAGTTCGCTACGCTCGATCTCTTCGGAATGCGATGGAGGATCGAT 51113
QY 964 atcgactcgctggaatccatccgcacaaagaaacacacccgcgcattgctatcataccccc 1023
Db 51114 CTTGATGCTGTGCGAGGCTCTTGCAGACGAAACACGCTTGTGCTTGTGTTATTAACCT 51173
QY 1024 acaaacctggcgagctgttactctctcagcaccatctctcaagg----- 1070
Db 51174 GGTAACTCTTGGCGGAATCTATAGCTACCGACATTTGATGAAGGTTAGCTTTAAACTT 51233
QY 1070 ----- 1070
Db 51234 TGCAATGAGGAACACAGACCAAAATTTGTTTGGAAATTTGATGATACAAATTTGTGT 51293
QY 1070 -----tcgcggaggtggcgaaaggtcggaaatattggtgattgctgacgaggtatc 1122
Db 51294 GTTTAGATTGCGGAATCGCGCAAAACATAGGGTTCTTGTGATTGCTGATGAGGTTTAC 51353
QY 1123 gcaacgctgtctggcgagcccgctcccaatcccaatggaggtgttgggcacatcacc 1182
Db 51354 GGTACTTCTGCTTTGGTAGCAACCGTTTGGCAATGGGTGTTGGATCTATGTT 51413
QY 1183 cctgtgctgccataggtctctgtcccaagtcatgtagtgcctggatggcggttga 1242
Db 51414 CTTGCTTACTCTTGGCTCTTTATCAAGAGATGATAGTTCCAGGTTGGCGACTCGGG 51473
QY 1243 tgggtgagcggtgaacaccccgaaagatctctacaggaactaagatctcta 1294
Db 51474 TGGTTGTCCACCACTGATCTTCTGCTGCTTCTTAAAGACCCCTAAGGTTAGTA 51525

RESULT 10
SC2H4/c 25970 bp DNA BCT 07-SEP-1998
LOCUS Streptomyces coelicolor cosmid 2H4.
DEFINITION AL031514
ACCESSION 93559956
NID
VERSION
KEYWORDS aminotransferase; beta-mannosidase; permease; secreted
beta-mannosidase; sugar transport system; sugar-binding
lipoprotein.
SOURCE Streptomyces coelicolor.
ORGANISM Streptomyces coelicolor
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 25970)
AUTHORS Oliver K. and Harris D.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 25970)
AUTHORS Parkhill, J., Barrall, B.G. and Rajandream, M.A.
JOURNAL Direct Submission
Submitted (01-SEP-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrall@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE 3 (bases 1 to 25970)
AUTHORS Redenbach, M., Kiese, H.M., Denapate, D., Eichner, A., Cullum, J.,
Kinashi, H. and Hopwood, D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
COMMENT Notes:
.Streptomyces coelicolor sequencing at The Sanger Centre is funded

by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2H4 lies between 265 and AH10 on the AseI-B genomic restriction fragment.
Location/Qualifiers
1. .25970
/organism="Streptomyces coelicolor"
/strain="A3(2)"
/db_xref="taxon:1902"
/clone="cosmid 2H4"
1. .1822
/gene="SC2H4.01"
<1. .1822
/gene="SC2H4.01"
/note="SC2H4.01, probable ATP/GTP binding protein, partial CDS, len: >606 aa; contains PS00017 ATP /GTP-binding site motif A (P-loop)"
/codon_start=2
/transl_table=11
/product="putative ATP/GTP binding protein"
/protein_id="CAA20595.1"
/db_xref="PID:e1319219"
/db_xref="PID:g3559957"
/db_xref="GI:3559957"
/db_xref="SPTREMBL:O86584"
/translation="IDSTAPKAPVITSLSPYQTCTADLCCKGGGAGTSFTFPNAA
DTDTITGYRWLLTTSKAEKAVTKGKVTSDVTPSLSGTOVLSEAKDVRSGAPAE
FSFVAPAGATGTHWEDDALPGSGVTIARDTAEGRHDAITLTAGAGNSTLARRGD
ADNSLWLDSDTITRAGIYAATAPAVNTRDSFTVSAYWITDTSQTRVMAAPGTG
SAFTIYASIKYKWFNRGADVKDPYLSLGDATAPLKVTHLAADFTRKDTN
KANDTIOLFVNGRPQFVLDGVTAYQPVWSSGGLQIGSLVGGWGEHFRGLDE
VNVQWVLTDEITQOALLEGVPANELVAOWDAASGTGVTKELSSYPAPSLTSLA
AGAVLDENNALVLDGSGAYATGVPVDEGTGFTVSARVOLDKAKLSPVGEALV
AGQASAGESSWALWAYKPADGYQWKFRTIYAVGADKVTQSAEAPGGDIATETWQ
VTVGDAQEPWEWTDPADSAKRTETRYGLHLVIGFDPQSPSTASGFTTTPQAGALAV
GTGSRGTTGHHLFGALEELRVMTGMSADQVRSQVLGGV"
188. .211
/gene="SC2H4.01"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
1949. .1953
/note="possible RBS upstream of SC2H4.02"
1959. .8510
/gene="SC2H4.02"
1959. .8510
/gene="SC2H4.02"
/note="SC2H4.02, unknown, len: 2183; very limited similarity to proteins of the RBS family e.g. WAPA_BACSU

FEATURES
source
gene
CDS
misc_feature
RBS
gene
CDS

Db 9569 GACGTACGGGTGCGGGTTTCCTCGCTCGGCTGGTGTGCTGACGGGGCGGAAGCAG 9512

RESULT 11
MSG81970CS/c
LOCUS MSGB1970CS 39399 bp DNA BCT 26-JUN-1996
DEFINITION Mycobacterium leprae cosmid B1970 DNA sequence.
ACCESSION L78815
NID g1377775
VERSION L78815.1 GI:1377775
KEYWORDS
SOURCE Mycobacterium leprae (clone: cosmid B1970) (tissue library: Lorist 6) DNA.

ORGANISM
Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.

REFERENCE
AUTHORS Elgmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.
TITLE Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae
JOURNAL Mol. Microbiol. 7 (2), 197-206 (1993)
MEDLINE 93188700
REFERENCE
AUTHORS 2 (bases 1 to 39399)
Smith, D.R., Richterich, P., Rubenfield, M., Butler, C., Lee, H.-M., Xu, Q., Gunderson, K., Chung, M., Maher, J.K., Deloughery, C., Aldrich, T., Imrich, J., Tulig, C., Smyth, A., Drilli, S., Avriuch, A.S., Rice, P., Abendschan, K., Aldredge, T., DeLoughery, C., Kirst, S., Safer, H., Connelly, S., McDougall, S., Elgmeier, K., Bergh, S., Cole, S., Robison, K., Jaehn, L., Gryan, G., Church, G.M. and Mao, J.
TITLE Prepublication submission
JOURNAL Unpublished (1996)
COMMENT This sequence was generated by the Genome Sequencing Center at Genome Therapeutics Corporation (Collaborative Research Division), Beaver St., Waltham, MA, 02154. Please contact Doug Smith (smith@eric.com) for further information. The sequence represents an insert of a Lorist 6 cosmid clone from a mapped set of clones constructed from M. leprae genomic DNA isolated from armadillo liver [3]. The sequence may not represent the entire cloned insert of the cosmid if an overlapping region was previously sequenced from another clone. Coding sequences larger than 60 amino acids were predicted on the basis of codon usage and homology information. An attempt was made to locate the most probable start site based on codon usage, homology, the presence of a Shine-Dalgarno sequence, or overlapping orf that suggested translational coupling. It is possible that the actual start site differs from the one selected.

FEATURES
Location/Qualifiers
1..39399
/organism="Mycobacterium leprae"
/specific_host="Dasypus novemcinctus"
/db_xref="taxon:1769"
/clone="cosmid B1970"
/tissue_lib="Lorist 6"
/note="The liver of the armadillo was used to isolate the Mycobacterium leprae."
BASE COUNT 7827 a 10529 c 12083 g 8960 t
ORIGIN

Query Match 5.8%; Score 111; DB 1; Length 39399;
Best Local Similarity 51.8%; Pred. No. 1.3e-05;
Matches 300; Conservative 0; Mismatches 275; Indels 4; Gaps 2;
Qy 689 gtaaccgccggtcggtccctcccccgcgcacgaagcgcgtggcagagcactgtcgc 748

Db 9552 GGTACTCCGACTCCAGGGCATCTCTGCCTGCACGTGCGGGTGGTTACCCGCTACGAGC 9493

QY 749 agggcggtccgtacatgtat---cggccgaacagctcttctcctcaccgcgcgggaccc 805

Db 9492 TGTCTCAGCGGTTCCCGCGATTTCGACGTGATGACGCTATCTATCTCGCAACGGGCTCTCG 9433

QY 806 agcgatcgaggtcataatacccggctgctggccagaccgcgcgcgaacattctgctcc 865

Db 9432 AGCTGATCAGATGACGTGCGAGGCCCTGTAGACAAACGGC-GATCCAAAGTCTGATCC 9374

QY 866 ccaggccaggctaccacaactacgagcgcgcccgcttcaacagcgtggagggtccggc 925

Db 9373 CCTCGCGGACTACCGCTGTGGACAGCGTCGACCTCGCTGGCCGTGGCACACCCGTAC 9314

QY 926 attcgacctatccccgcacaagggtgggagatgacatcgactcgctgggaatccatcg 985

Db 9313 ATTACCTCTCGCAGCAACCAAGGCTGGCAACCATATCGCGACGTGGAATCTAAGA 9254

QY 986 cgcacaagaacaccaccgcctatggtcatataaaccccccaaccccgctgcgcagcggtt 1045

Db 9253 TCACCGAGCGCACCAAGCACTGGTGTGATCAACCCCAACCCCTACCGGCGGCT 9194

QY 1046 actctacgacctctgtccaaaggtcgcgaggtggcgaaaaggtcggaatttggtga 1105

Db 9193 ACAGCAACGAAATCCTTAACCAAAATAGTCGACCTAGCAGCTAAGCATCAGCTGCTGC 9134

QY 1106 ttgctgacaggtatatacgcagagtggttctctggcagcgcccccttcccaatgggag 1165

Db 9133 TTCCCGACGAGATACGCAAGAATCTCTATGACGACACCAAGCATCAGCTGGGCT 9074

QY 1166 tgttgggcacatcacccctgtctgctccataggtctctgtccaaagtcagtagtagc 1225

Db 9073 CAATTCGCGCTGATCTACGTGCTGACCTTCAACGGCTTTTCGAAAGCTACCGTGC 9014

QY 1226 ctggatggcggttgatgggtgtagcggtgtagcagcccca 1264

Db 9013 CCGGGTACCGGTCCGGCTGGCTAGCGATCACCGGGCCCA 8975

RESULT 12
AB018112 70475 bp DNA PLN 20-FEB-1999
LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MAB16,
DEFINITION complete sequence.
ACCESSION AB018112
NID 93702730
VERSION AB018112.1 GI:3702730
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:MAB16.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (sites)
Nakamura, Y.
Structural Analysis of Arabidopsis thaliana Chromosome 5. IX
Unpublished (1998)
2 (bases 1 to 70475)
Nakamura, Y.
Direct Submission
Submitted (06-OCT-1998) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail:ynakamura@kazusa.or.jp, Tel:+81-438-52-3935,
Fax:+81-438-52-3934)
Location/Qualifiers
1..70475
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"

FEATURES
source

[illegible]


```
KNRQNLPGYVLVRLMDLTNESWGVVNTPGVTGFGVGNAYDPYPLTLDLDEIVKMLAPEA
QEKAAKAAEAGLPAAPAVKRTIEVLDFEVDGSDTVYVDGFPATLQATINEINPDSKKV
KGLVEIFGRTPEVLSFDIOQKN"
4373..4807
/gene="rplK"
CDS
4373..4807
/gene="rplK"
/note="Accession number of rplK has been submitted as
D90512"
/codon_start=1
/transl_table=11
/number=1
/evidence=not_experimental
/product="ribosomal protein L11"
/protein_id="BAA09303.1"
/db_xref="PID:d1009943"
/db_xref="PID:g1217912"
/db_xref="GI:1217912"
/transl_table="MPPKKKVTGLIKLQIKAGAAANPAPVPGPALGQHGYNIMEFCKA
YNAQTSQRMVVPVEITVYDDRSFTFTKTPPAARLILKHAGIEKSGSGEPHKTQVAK
LTAQVKEIAELKMPDLNANDIDAAVKIIAGTARSMGVTVEG"
4896..5624
/gene="rplA"
4896..5624
/gene="rplA"
/codon_start=1
/transl_table=11
/number=1
/evidence=not_experimental
/product="ribosomal protein L1"
/protein_id="BAA09303.1"
/db_xref="PID:d1009944"
/db_xref="PID:g1217913"
/db_xref="GI:1217913"
/transl_table="MKRSKTLRAADAKVDEKLYALEAVRLAKETSTTKFDGTVEVA
FRIGVPRKADQVRGTYNLPHGTGTARVLVATGDRAAAEAAGADIVDDDELIN
IAGNRLNEDFADVLDLGMKGRVLRGLVPGRLMPNPKTGTVMDDAKAVTEIKGG
KIEFRVDSKSNLHFTIIGKVSFTDEQVENVYGAALDEILRLKPSAAKGRYIKKAALSTT
MGPIHLDNRTNRLNLLVEEDPAV"
6159..6719
/gene="rplJ"
6159..6719
/gene="rplJ"
/codon_start=1
/transl_table=11
/number=1
/evidence=not_experimental
/product="ribosomal protein L10"
/protein_id="BAA09304.1"
/db_xref="PID:d1009945"
/db_xref="PID:g1217914"
/db_xref="GI:1217914"
/transl_table="MFTPNKAAVAELKDAFOSSNAAVLTYEYRGLTVAOLKTLRLSLG
ENAYAVVNTLTAKAANQAGITALDEHFGATVAFITGDPVESAKSLRDFAKDNP
LIKAGVLGSKALTADKIKKLADLESREVLKLAGAFKQKQSAASLFOALPSKFVR
TAKRFASSPSRAVPSNSARALIHA"
6781..7182
/gene="rplL"
6781..7182
/gene="rplL"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ribosomal protein L12"
/protein_id="BAA09305.1"
/db_xref="PID:d1009946"
/db_xref="PID:g1217914"
/db_xref="GI:1217914"
/transl_table="MEGSPIMAKLSQDLDLAQFEETLIELSEFVKAEFEKFDVTA
AVAYAGPVAGVAPRAEEQDEFVILTGAGDKKIQVIRKVVRELTSLGLKEAKDLVDG
APKPVLEKVAEAKAEASLKAAGAAVEK"
1243 a 2483 c 2560 g 1123 t
```

```
Query Match 5.4%; Score 103.6; DB 1; Length 7409;
Best Local Similarity 48.5%; Pred. No. 0.00014;
Matches 317; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

QY 543 caagatcagcgagcgagcgagcagagagagggcgcccgctgctgcgctgccccagc 602
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2502 CACCTCGCGGTGGACCAAGGCCAAGGCCCTCAAGGCCGCCGCCGGTGTATCGG 2443
QY 603 ggaccgcgttcggttcggttcggttcggttcggttcggttcggttcggttcggttc 662
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2442 CTTGGTGGCGGGGAGCCGACACTTCCGACCCCGGACTATACATCGAGCGGGGTCTG 2383
QY 663 cgcgcgtgcacggcgccagttcaactgtacctccgcgcgcgcgcgcgcgcgcgcgcgcgc 722
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2382 GGCCTGCCGAACCCCAAGTACACCGCTACGCGCGCGCGGGGTCTGCCCGAGCTCAA 2323
QY 723 aagcgcgtgcagagcacctgtgcagggcggtgcgcgtacatgctatcgcgcgcgcgcgc 782
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2322 GCGCGCATCGCGCCCAAGACCCGTGCGGACTCCGGCTACGAGGTGAGGCTCGCAGGT 2263
QY 783 ctctccaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 842
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2262 CTTGGTGACCAACGGTGGCAAGCAGCGGATCTAGAGGCCCTTCGGGCCAT---CCTGGA 2206
QY 843 cgcgcgcgcacattctgtctcccgagggcgaggtacctcccaaaactacgagcgcgcgcg 902
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2205 CCCGGGTGACGAGGTCTATCGTCCGCTCGTACTGTGACCACTACCGCGAGTCCATCCG 2146
QY 903 gttaacagcgctgaggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 962
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2145 TCTCGCGCGGGTGTCCGGTTCGATGTGCGCGGAGAGACCACCGGTACCGCGGTCTC 2086
QY 963 catcgactcgctggaattccatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1022
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2085 CGTCGAGCAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026
QY 1023 caaacaaccgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1082
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2025 GTCCAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
QY 1083 gaaaagcgtcggaatttggtgattgctgacgaggtatcgcaagctggttctggtgcag 1142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1965 CGCGGAGCAGCGCTGTGGTGTGTCGACGACGAGATCTAGAGCACCCTGGTCTACGGCGGA 1906
QY 1143 cgcgcgcgttcacccatggaggttttgggcacatcacccctgtgtcttccat 1196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1905 GCGGAAGTTTCACTCGCTCGCGTCTCTGTCGCCGCCCTCGCGCGCAAGTGCAT 1852
```

RESULT 14
MTCY279/c

LOCUS MTCY279 9150 bp DNA BCT 17-JUN-1998
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 17/162.
ACCESSION Z97991 AL123456
NID 93261837

VERSION 297991.1 GI:3261837

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.

REFERENCE 1 (bases 1 to 9150)

AUTHORS

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekoa, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and

Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov 12;396(6707):190]]
2 (bases 1 to 9150)
Direct Submission
Parkhill,J.
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2276330.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of rB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

Location/Qualifiers
1..9150
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
<1..9106
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y279"
82..86
/note="possible RBS upstream of Rv0336"
90..1601
/gene="Rv0336"
90..1601
/gene="Rv0336"
/note="Rv0336, (MTCY279.03), len: 503 aa; almost identical to M. tuberculosis hypothetical protein MTCY20610.05 (99.8% identity in 503 aa overlap) possibly due to a recent gene duplication, also similar to other M. tuberculosis hypothetical proteins eg. MTCY049.05 (61.1% identity in 347 aa overlap) and Y140_MTCY0 Q10709 hypothetical 58.9 kd protein cy49.40 (550 aa), fasta scores: Opt: 676 z-score: 978.7 E():0.41.7% identity in 513 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0336"
/protein_id="CAB10791.1"
/db_xref="PID:e330320"
/db_xref="PID:e330320"
/db_xref="GI:2276333"
/db_xref="GI:2276333"
/transl_table=11
/translation="MPSPEIAHDFEFCAPRTTRVSAAFIDRICSATRAENRAAA AQLVALGELFAIRMSRGGREWMYDVAAREVAALRIISQGLAASRLRYARAME RLPTKAVFSAGDYGILMEATIVYRTDILVDPDLAAVDLAQLAANVARWPSMTKARLA GONDKIVARADADAVRRKEYOQORQWGESODGVCQIGSLAVDAHALDARLSAL AGTVCEDHPSREORADALGALAGADRLGCGCGADCAAGKRAPAPPVYTHLIAEA ATINGTSAPASQNNADLITAEILVAELAKTATLVLVHPGDAPPEPGYAPSKALADF VCRDLTCRWFQDNPCLDHTIFYAAGGPTTHASNLKCYCRTHLVLTFWGRDQQLPDGTLILTSPSGHTVYSTVPSGALLFSLCHFSFGIPAPEADPPYDHCQRTAMFKA

RRRTAQDRAYRIATERRONHARQRAQVLTOTAAATDTHGPPPDNDPPPF"
complement(1771..3060)
/gene="aspC"
complement(1771..3060)
/gene="aspC"
/note="Rv0337c, (MTCY279.04c), len: 429, aspC, aspartate aminotransferase, similar to many eg. TR:Q48143 (EMBL:U32823) aspartate aminotransferase (404 aa), fasta scores: Opt: 1646 z-score: 2132.9 E(): 0, 57.2% identity in 404 aa overlap; also some similarity to M. tuberculosis protein MTCY06G11.12 (27.2% identity in 383 aa overlap)"
/codon_start=1
/transl_table=11
/product="aspC"
/protein_id="CAB10792.1"
/db_xref="PID:e330321"
/db_xref="PID:g2276334"
/db_xref="GI:2276334"
/db_xref="SPTREMBL:O33267"
/translation="MDNDGTIVDTTQHPWHTASHORAFQAQSAKLQDLVYEIRGP VHQAARLAEGRHILKLNIGNPAFGFEAPDIVMRDIIQALPYAQCYSQOILSAR RAVTVYELVPGPRFDVDDVYLGNGVSELTITLQALLDNGQVLIIPSDYPLMTAS TSLAGCTPVHYLDQETQGPQDIDLESKITERTKALVINPNPTGAVYSCETLQOM VDAARKHQLLLADEIDYKLIYDDAKHISLASIAPMCLITFNGLSKAYNAVYRAGW LATIGPKHASSFIEGILLANMRLCPNVPQAIOVALGHQSIEDLVLPGRLLQEQ RDTAWTKLNEIPGVSCVPAGALYAFRLDPEYDIDDDQLVDLLSLKILVTQGT GFNWPADHLRLVTLPSWRDLAAAIERLGNFLVSRQ"
complement(3091..5739)
/gene="Rv0338c"
complement(3091..5739)
/gene="Rv0338c"
/note="Rv0338c, (MTCY279.05c), len: 882, unknown iron-sulfur protein, first of several possible start sites chosen. Some similarity to other hypothetical proteins eg. YWJF_BACSU P45866 hypothetical 79.2 kd protein in acda (705 aa), fasta scores: Opt: 1188 z-score: 894.6 E(): 0, 34.8% identity in 682 aa overlap. Contains PS00070 Aldehyde dehydrogenase:cysteine active site and two of PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0338c"
/protein_id="CAB10793.1"
/db_xref="PID:e330322"
/db_xref="PID:g2276335"
/db_xref="GI:2276335"
/db_xref="SPTREMBL:O33268"
/translation="MTQTLIRLILGMSMTAVGVFALRRVWMLYKLVMSGQPASGRT DNLGTRITQISEVLGQRLKWSIFGLAHFTTMMGFFILLTVIYAYGLLFEERFHI PVTRWDALGLODFATVFLGITTFAIRILRNPREIGRSSRFYSGHNGAWLVLL MIVNVIWTVYLVRSVAVNGTLPYNGAFSLQSLFGLRPLGQPANIEITETALLHI GVMNAELLVHSHKHLHFLAPINVTFKLPDGLPPLPLEADGKPLDFENPSDAVE GRKIDFTWKGMDLATCECGRCSQCQPAWNTGKPLSLKLVINDLRDHWMAKAPLI LGQDASAGGAGHGHVHPSGFGVPGHPEQATRLVLTETEEQGVIDDPVWMSQ VTCGACVEQCPDIEHYDTHVIMRRYQWMMSEFSELSVLFKNLTKGPNQNASD RTNIVDEVDVPVYQGVDFDGYEYLFVWGCAGAYDDAKKTTKAVAEILAVARYK VLVLAGETCNGDSARSGNEFLFOOLAQOAVETLDGLFEGVETVDRKIVVTCPCFN TICKEYRQLGANYTVLHHTOLLNRLVRDKRLVPTVPVSODITYVHPCYLGHNKAYEA PRELIGAASLTEMPRHADRSFCCGGAGGAMWEEHIGRINHERVDEALDATAI ATACPFRCVMVTDGVNDREBEASGEVLDVAQLGLSLDHDKAUPAKGTAAKQAI ERAPKAAPKAAAPVTPVEAPAEAPQAPAPAPAPVKGLGMAAGAKRPPGAKAPTA APAAPAPVKGLGTAAGAKRPPGAKTTPPPAPGLAEPAAQPEAKPQPEAPAPKPT DGDPAAPAPVKGLGTARGARPPGKR"
complement(4174..4209)
/gene="Rv0338c"
/note="PS00070 Aldehyde dehydrogenases cysteine active site"
complement(4531..4566)
/gene="Rv0338c"
/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature"
complement(4822..4857)
/gene="Rv0338c"

misc_feature

misc_feature

misc_feature


```
CDS
/ gene="stgR"
2219..3202
/ gene="stgR"
/ note="lysR-type transcriptional regulator"
/ codon_start=1
/ transl_table=11
/ product="StgR protein"
/ protein_id="CAA06437.1"
/ db_xref="PID:ei1318409"
/ db_xref="PID:g3550624"
/ db_xref="GI:3550624"
/ db_xref="SPTREMBL:O88132"
/ translation="MDLDLALVRSFAITAEELHFRRAERLFTIOGLSORIRLELL
LGVSFLVGNRSVALTEGQFLAPARRVMEADAADAAVQVPRRRPLRVVDGRLA
PLQVLRKLEEDPELPKLSMROSLAALPALARDEIDTAFGRVHDVQWNAEFSHRL
VRLEPLRALVPRTHPLAGRDVRIDELRSCLWQPRNGVPSEWEHYVGRADSFGLSL
AFSGAALSDHEFLHLSHGEMACIAGADVSTAFSPDIRSIPLDVPTFPYPMWVWRR
QAGHPLVDRLVGLAKRSAGDMLTHVPGEIWLPAIDRALLRGAGVDVDAVRG"
BASE COUNT 447 a 1112 c 1189 g 492 t
ORIGIN

Query Match 5.4%; Score 102.2; DB 1; Length 3240;
Best Local Similarity 50.9%; Pred. No. 0.00024;
Matches 301; Conservative 0; Mismatches 278; Indels 12; Gaps 2;

QY 679 cagttcaactgtatcccccgccggtcggcctccccccgcacgaagcgcgctggcgag 738
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1202 CAGGCGCAGCGGTACACGACTCGCGCGCGGTCCGTGTCGCGCGCGCGCGCGCGCCG 1143

QY 739 cactgtcgaggcggtgcgtacatgctatgcccgcgacgacttctctaccgccggc 798
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1142 CGTATACGCGCTCGGCTCGACGAGGTCTCGGTGGACGCGTCTTCTCGGCAACGCG 1083

QY 799 gggaccagcgatcgagtcataatcccggtgctggcccgacacgcgcggcccaacatt 858
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1082 GTGTCGAGCTGTCAGATGCGGTGTCAGGCGCT---CCTGGAGACGCGGACGCAATC 1026

QY 859 ctgtccccagcggtaccacaaactacgagcgccgcttcaacaggctggag 918
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 CTCATCCCGCGCGGACTTCCGCTGTGGACCGGGTGACACCTCTCCGAGGCAAG 966

QY 919 gtccggcatttgacactatccccacaaaggggtgggagatcgacatcgactcgtggaa 978
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 GCCGTCCACTAGCTGCGACGAGCGCGGACTGTGTACCCGACCTCGACGACATGGCC 906

QY 979 tccatgcgcgaagaacacaccgccatggtcatcataaaccacacaccccgctgcggc 1038
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 TCGAAGATCACCACACGACCAAGCCGTCGAGATCATCAACCCCA-----CGGGC 855

QY 1039 agcgtttactctacgacctgtccaaggtcgcgagggtggcgaaaggctcggaata 1098
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 GGGGTCTATCCGAGAGATCTGGAGGCACTCTCGATCTCGCGCGGCGGACGCGCTG 795

QY 1099 ttgggtattgtcgagggtatcacggaagctggttctgggagcgcccgcttcacca 1158
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 794 ATGGGTCTTCGCGGACGAGATCTACGACCATCTGTACGAGCGCGCTGCACACACAG 735

QY 1159 atgggagtggttgggcacatcacccctgtgctccatagggtctctgtccaagtcagg 1218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 734 GTGGCGGCGCTCCCGCGGATCTGTGGTCCTACCTTCTGCGGTCTTTCGAAGACGTAC 675

QY 1219 atagtgcctggatggcggttgatgggttagcggtgtacgaccccgagaaag 1269
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 CGCGTGGCGGGTTCCGTTGCGGGCTGGTGGTCATCACCGGCGCGCGCGGAG 624
```

Search completed: October 1, 1999, 15:34:36
Job time: 6179 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
C	1	1910	100.0	1910	1	V48148
	2	1146.4	60.0	1660	1	V48147
	3	95.4	5.0	390	1	Q21833
	4	95.4	5.0	390	1	Q36859
	5	80.2	4.2	985	1	V44439
	6	80.2	4.2	985	1	V64548
	7	77.8	4.1	3946	1	T93610
	8	75.4	3.9	2539	1	T61454
	9	75.4	3.9	2809	1	T61455
	10	74.2	3.9	2277	1	V05372
C	11	74.2	3.9	2277	1	V13836
	12	73	3.8	114955	1	X53491
	13	69.6	3.6	114955	1	X53491
	14	69.4	3.6	3946	1	T93610
	15	67.4	3.5	44377	1	T78508
	16	67.4	3.5	44377	1	T80414
	17	67.2	3.5	1439	1	V44424
	18	67.2	3.5	1439	1	V64533
	19	66.6	3.5	15872	1	T68715
	20	65.8	3.4	1620	1	Q22482
C	21	65.8	3.4	2167	1	Q22484
	22	65.8	3.4	2668	1	Q22485
	23	65.2	3.4	2742	1	V49291
	24	65.2	3.4	1166	1	X34248
	25	65.2	3.4	1166	1	X34249
	26	65.2	3.4	1164	1	X34250
	27	64.2	3.4	2198	1	T71711
	28	63.8	3.3	24379	1	T93095
	29	63.8	3.3	24379	1	V25925
	30	63.2	3.3	494	1	V44433
C	31	63.2	3.3	494	1	V64542
	32	62.8	3.3	1272	1	X09010
	33	62.6	3.3	15672	1	Q10613
	34	62.6	3.3	30001	1	T61016
	35	62.6	3.3	985	1	V44339
	36	62.6	3.3	985	1	V64548
	37	62.6	3.3	30001	1	X05110
	38	62.2	3.3	538	1	V44437
	39	62.2	3.3	538	1	V64546
	40	62	3.2	1320	1	Q22481
C	41	61.8	3.2	29879	1	Q46806
	42	61.8	3.2	2888	1	T93268
	43	61.2	3.2	12001	1	T76213

Db	421		AGCAACGGGCATCGCGCGCGGCAGAGGAGGAGGCGGTGGAGTGGAATTCGCG	480
Qy	481		ggTgcAAAgagcgTgctgTgcggcgagcggggcgaaatagcatcgccggcgatacgg	540
Db	481		GGTGCAAAGGACGGCTGCTGCGGCGCAGCGGGCGCAACATGAGCATCCGGGCGCATCGG	540
Qy	541		tacaaatcagcgcgagcgTgcagagagaagggcgcgccgTgctgcgcTgcccac	600
Db	541		TACAAGATCAGCGCAGCGTGCAGAGAAAGGGCGCGCGCCGTGCTCCGCGTGCGCCAC	600
Qy	601		ggggaccgTccgTgtcccgccTccgcacgcgTgcgagggcgagcgagcgTgcgc	660
Db	601		GGGACCCGTCGTGTCCCGGCCCTCCGCACGGCGCTCGAGGCGGAGGACCGGTGCGCC	660
Qy	661		gcccgcgtgogaacggccagTtcaactgtaccgcccgcgTcgggctcccgcgcgca	720
Db	661		GCGCGCGTGCACACCGCCAGTTCAACTGCTACCCCGCGCGCTCGGCTCCCGCGCGCA	720
Qy	721		cgaagcgcgTggcagacgacacTgTcgaggcgTgcgTatcgtatcgtatcgcgacgac	780
Db	721		CGAAGGCCGTGCAGAGCACTGTGCGAGGCGTGGCGTACATGCTATCGGCGCGACAC	780
Qy	781		gtcttctcacgcggcgggagccagcgatcgaggttcaatacccgTgctgTgcgcag	840
Db	781		GTCTTCTCACCGCGGGGACCCAGGCGATCGAGGTCTAATATCCCGGTGCTGSGCCAG	840
Qy	841		acgcgcggcgcaaaattgtgtctcccaaggcgagTgtaccaaactacgagcgcgcc	900
Db	841		ACGCGCGGGCGCAACATCTGTCTCCAGGCGAGGCTACCCAAACTAGAGGCGCGCGCC	900
Qy	901		gegtTcaacaggtgagTgcggcatttcgacctcatcccgacaaagggtggagatc	960
Db	901		GCCTTCAACAGGCTGGAGGTCCGGCATTTTCGACCTATCTCCCGCACAGGGGTGGGAGATC	960
Qy	961		gacatcgactgctggaattccatgcgcgacaagaacacacgcgcattgtcatcataac	1020
Db	961		GACATCGACTCGTGTGAATTCATCGCGCACAAAGAACACCGCATGGTTCATCATAAAC	1020
Qy	1021		cccaacaacccTgcggcagctTtactctcagaccatctgtccaagTgcgaggtg	1080
Db	1021		CCCAACAACCCGTGCGGACGCTTTACTCTACGACACATCTGTCTCAAGTCTCGGAGGTG	1080
Qy	1081		gcgaaaggtcggaattatTgTgattgTgcgagaggtatacggcaagctggttctgggc	1140
Db	1081		GCgAAAGGCTCGGAATATTGGTGATTGCTGACGAGGTATACGGCAAGCTGTTCTGGC	1140
Qy	1141		aggccccgttcatcccaatggagTgTtTgggacatcaaccctgtgctgtccatagg	1200
Db	1141		AGCGCCCCGTTCATCCCAATGGAGTGTGTGGGCACATCACCCCTGTGCTGTCCATAGG	1200
Qy	1201		tctctgtccaaTcatggatagTgcTggatggcggtTggatgggtagcggtgtacgac	1260
Db	1201		TCCTCTGTCAAAGTCATGGATAGTGCTGGATGGCGGCTTGGATGGGTAGCGGTACGAC	1260
Qy	1261		ccagaaagatcttacagaaactaagatctctcatcaaatcaagaaattacctcaatgtc	1320
Db	1261		CCAGAAAGATCTTACAGGAACCTAAGATCTCTACATCAATTAGGAATTACCTCAATGTC	1320
Qy	1321		tcgacagaccagcaaccttcatcgagcgactctctcgtatcttcgattctgagaacaaag	1380
Db	1321		TCGACAGACCCAGCAACCTTCATTTCAGGCAGGCTCTTCTCTAGATTCCTTGAGAACAAAG	1380
Qy	1381		gaagattctttaagcgattattggTgcttaagggaatcatcagagatatgtacaaa	1440
Db	1381		GAAGATTCTTTAAGCGGATATTGGTCTGCTFAAAGGAATCATCAGAGATATGCTACAAA	1440
Qy	1441		caataaagaaacaaatcacattacatgtctctcaaacgacgaagatcaattgttc	1500
Db	1441		CAAAATAAGGAAACAAATACAAATACATTCCTCTCAAGCCAGGAAGATCAATGTTGTC	1500
Qy	1501		atggtgaaactgaacttacctcttttggaggaaatagacgatgacattgattttgtc	1560

Db	1501	ATGGTGAACCTGAACCTACATCTTTGGAGGAATAGACGATGACATGATTTTCTGCTGC	1560
Qy	1561	aagctcgcaaaagaagaatcagtaaatcttatgccagggagtgcttcttggaaatggcaaac	1620
Db	1561	AAGCTCGCAAAAGAAGAATCAGTAATCTATGCCAGGAGCTGTTCTTGGAATGGCAAC	1620
Qy	1621	tgsgtcgcgaattacttttgccttgcttgctccattctcttcgaagatgctcgcgaaggatc	1680
Db	1621	TGSGTCGCGAATTACTTTTGCCTTGTCTTGCATCTTCTTCAAGATGCTCTCGGAAGGATC	1680
Qy	1681	aaatcattctgtcaaaaggaacaagaagaattcgacgatgattcctagtgttatctt	1740
Db	1681	AAATCAATCTGTCAAAAGAACCAAGAGAGAATTCGACGATGATTGCTAGTGTATATC	1740
Qy	1741	tgactgaagctgtaaatcatctcccgagtcgcccatcttatcttccaataaaatggaaact	1800
Db	1741	TGACTGAAGCTGTAAATCATTTCCAGTATCCCATCTATATCTTTCAATAAATGGAAC	1800
Qy	1801	tttagttctctatgaataagaatcaaatctctcttgaaatgtctctgtgtgtgtggcct	1860
Db	1801	TTTAGTTCCTCTATGAATAGAACTCAACATCTCCTTGATATGTTCTGCTGTTGTGGCCT	1860
Qy	1861	ggacgaaacatgtagaattgtatgttagtgaagttaaaaaataaaaaa	1910
Db	1861	GGACGAAACATAGTGAATGTTAGTAGGAAGTTAAAAAATAAAAAA	1910
RESULT 2			
V48147			
ID	V48147 standard; cDNA to mRNA; 1660 BP.		
AC	V48147;		
DT	27-OCT-1998 (first entry)		
DE	Nicotianamine aminotransferase 49564.15 molecular weight protein, gene		
KW	ds; nicotianamine aminotransferase; plant; iron absorption;		
KS	iron deficiency chlorosis.		
OS	Gramineae sp.		
PH	Location/Qualifiers		
Key	62..1447		
FT	CDS		
FT	/*tag= a		
FT	/product= "Nicotianamine aminotransferase"		
FT	EP-8650499-A2.		
PN	26-AUG-1998.		
PD	19-FEB-1998; 102891.		
PF	21-FEB-1997; JP-037499.		
PR	(SUMO) SUMITOMO CHEM CO LTD.		
PA	Mori S, Nakanishi H, Takahashi M;		
PI	WPI; 98-439341/38.		
DR	P-PSDB: W61642.		
DR	New nicotianamine aminotransferase protein and DNA - useful for		
PT	enhancing iron absorption of plant cells		
PS	Claim 4; Page 12-13; 17pp; English.		
CC	The nicotianamine aminotransferase can be used in a plasmid to transform		
CC	plant cells to produce cells with enhanced iron absorption, and it is		
CC	implied (though not stated) that plants with improved resistance to iron		
CC	deficiency chlorosis in calcareous soils can be regenerated from the		
CC	transformed cells. The gene fragment can be used to detect, amplify		
CC	and/or isolate nicotianamine aminotransferase genes.		
SQ	Sequence 1660 BP; 423 A; 442 C; 430 G; 365T;		

[illegible]


```
QY 255 gaacggcaagagcaacggccatgccagggccactgcgaacggcgaacggcagggccactgc 314
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 TGGCGCGCGGCACAGG---TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
QY 315 gaacggcaagacaaacggccacgcgcagagagcaacggccatgctgagggccgcgcgcgaa 374
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 TACCGGGCGGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 299
QY 375 cggcgagagcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 359
QY 428 ggcacggcgggcgagagagagagagagagagagagagagagagagagagagagagagagag 487
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TGACGGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 419
QY 488 aggcacggcggtgctgagagagagagagagagagagagagagagagagagagagagagag 547
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
QY 548 tcagcgagagcgtgcagggagagagagagagagagagagagagagagagagagagagagag 607
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 AGGTCTCGCGACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
QY 608 cgtcgtgttcccgcttcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 667
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 ACGCGCGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
QY 668 tgcgcacggcagctcaactgtaccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 708
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 ATGGCGGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639

RESULT 6
V64548
ID V64548 standard; DNA; 985 BP.
AC V64548.
DT 27-JAN-1999 (first entry)
DE M. tuberculosis immunogenic polypeptide XP25 3'-end DNA.
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis; ss.
OS Mycobacterium tuberculosis.
PN W09816646-A2.
PD 23-APR-1998.
PF 07-OCT-1997; U18293.
PR 13-MAR-1997; US-818112.
PR 11-OCT-1996; US-730510.
PA (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI Reed SG, Skeiky IAW, Twardzik DR, Vedvick TS;
DR WPI; 98-261042/23.
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
PS Claim 4; Page 176; 230pp; English.
CC This sequence encodes an immunogenic portion of a soluble Mycobacterium
CC tuberculosis (MT) antigen which can be used in a method for inducing
CC protective immunity against tuberculosis (TB). This sequence can be
CC formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T;

Query Match 4.2%; Score 80.2; DB 1; Length 985;
Best Local Similarity 48.7%; Pred. No. 8.9e-07;
Matches 312; Conservative 0; Mismatches 318; Indels 11; Gaps 3;

QY 75 catggccaccgtacgcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 134
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 CAGCGGTACCGGTGGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62
QY 135 cgcgaacggcagagcaacggccatgctggtgctggtgctggtgctggtgctggtgctggtgct 194
```

```
Db 63 CAACACGGGTGTGGCGCGGCACCAACCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
QY 195 ccatggcgttgatgccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 254
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 CGCACCGGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
QY 255 gaacggcaagacaaacggccatgctgagggccactgcgaacggcgcgcgcgcgcgcgcgcgc 314
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
QY 315 gaacggcaagacaaacggccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 374
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 TACCGCGCGCACCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 299
QY 375 cggcgagagcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CGCGCGCGCGGTGACGGCGCGCATGGGCGCGAGCGGTCTCGGCGCTCCGCGCTTCCGCGCTT 359
QY 428 ggcacggcgggcgcgagagagagagagagagagagagagagagagagagagagagagagagag 487
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 TGACGGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATCA 419
QY 488 aggcacggcggtgctgagagagagagagagagagagagagagagagagagagagagagagag 547
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
QY 548 tcagcgagagcgtgcagggagagagagagagagagagagagagagagagagagagagagagag 607
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 AGGTCTCGCGACACCGCGCGCGGTTCGCGCGGTGACGGTGGGCGCGGTGGCGC-CGCCGGCA 538
QY 608 cgtcgtgttcccgcttcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 667
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 ACGCGCGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
QY 668 tgcgcacggcagctcaactgtaccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 708
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 ATGGCGGCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639

RESULT 7
T93610/c
ID T93610 standard; DNA; 3946 BP.
AC T93610;
DT 27-APR-1998 (first entry)
DE Mycobacterium tuberculosis genomic DNA fragment (I).
KW Tuberculosis; mycobacteria; infection; diagnosis;
KW antimycobacterial; antibiotic; vaccine; ss.
OS Mycobacterium tuberculosis.
PN W09741252-A2.
PD 06-NOV-1997.
PF 18-APR-1997; E01973.
PR 29-APR-1996; DE-017184.
PA (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
PI Espitia C, Honisch C, Moreno C, Singh M;
PI WPI; 97-549750/50.
DR P-PSDB; W31852-55.
PT New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
PS Claim 1; Fig 9; 55pp; English.
CC New claimed DNA (A) has one of 3 isolated Mycobacterium
CC tuberculosis DNA sequences of 3946 bp (I) (T93610), 2653 bp (VI)
CC (see T93611) or 440 bp (IX) (see T93612), or hybridises to, or is a
CC fragment of these sequences. Also claimed are RNA transcripts
CC of (A), and proteins encoded by (A). To isolate (I), a cosmid
CC library of M. tuberculosis DNA was screened with degenerate probes
CC (see T93621) containing GC-rich regions and designed to encode part
CC of a proline-rich protein. (I) contains open reading frames (ORFs)
CC for 4 proteins (see W31852-55). These ORFs can be amplified and
CC cloned into vectors for expression in E. coli or other hosts.
CC Clones (VI) and (IX) were obtained from a phage library and
CC together encode 3 proteins (see W31851 and W31856-57). The claimed
```

DNA sequences can be used for diagnosing tuberculosis and other mycobacterial infections in humans or animals, for identifying mycobacteria in (clinical) samples by hybridisation or amplification, including differentiation between strains, as well as for epidemiological studies, for monitoring vaccination or for the development of anti-mycobacterial drugs and vaccines.

Sequence 3945 BP; 518 A; 1541 C; 1345 G; 542 T;

Query Match 4.1%; Score 77.8; DB 1; Length 3946;
Best Local Similarity 47.8%; Pred. No. 3.1e-06;
Matches 318; Conservative 0; Mismatches 342; Indels 5; Gaps 3;

QY 77 tggccaccgtacgccagagcaggagtgcgccaagcgcttgccgtgcccgaacgcg 136
|||||
DB 3584 TCGCCACGAGTGGGATCGCGGGTGTCGGCGGAGCCGGCGGCACGCCATGTGTTCCGCGC 3525

QY 137 cgaacgcaagacaacggccatggctggctgcgcgtgaaacggtaagaagcaaacgccc 196
|||||
DB 3524 CCGCGCGCGCGCGCGCGCGCGGCGGCGCTAGTAGCCCTCACGGCGGGCGCGCGCGC 3465

QY 197 atggcgtgatgcgcagcgcgaacgcgaagacaacggccatggctggctgcccgaacgcga 256
|||||
DB 3464 CCGCGCGGGGTTCGCGCAACCGCGCTGCTCTTCGGCGCGCGCGAGTCTCGCGCGGGCGCG 3405

QY 257 acggcaagagcaacggccatccgagggccaactgcgaacggcgcagggccactgcga 316
|||||
DB 3404 CGGATTTCACGAACGGCAGTGCATCTCGCGGGCGGGCGCGCGCGCGCGCGCGCGC 3345

QY 317 acggcaagaccaacggccaccgcgagacaacggccatgctgagccgccgcagcgcgaacg 376
|||||
DB 3344 TGTTTCGCACCACGGCGGGGTCCGGCGGATCCGTTGGGGCCGGCAGACTCCGGCGGGCGCGGTG 3285

QY 377 gcaagagcaacgcatgcgcgagactccg-cggcgaacgcgcagagcaacgcggcatgcg 435
|||||
DB 3284 GGGCGGTGGCGCTGGCGGCTGTCTCGCGCGCGCGCGCACCGCGGGCATGGCGGGTTCG 3225

QY 436 gcggcgcgcgcagagggagggcggtgagttggaatttcgcgggtgccaaggacggc 495
|||||
DB 3224 CCGACAGCAGCTTCGGCGGAGTCGCGGGCGCGGGCGCGGGCGGTGTTCCGGCGCAG 3165

QY 496 gtgctgcgcgcgaacggggcgaaaataagcatcgcgggcgatacggtaacaagatacgcgcg 555
|||||
DB 3164 GCGGGGAAGCGCGCAGCGCGGCACACGCCCTGCTGCCGCGGGGACCGCGGGCGCGCGCG 3105

QY 556 agcgtgcagaggaaggcgccgcgtgctgcgcgtggccccacgcgggaccgcgttccgtg 615
|||||
DB 3104 CCAATGCCGG--CATGCTCGCCCTCGCGCGCGCGGGCGCGCGCGAATCGTGGGG 3048

QY 616 ttcccgcccttcgcacgcgcgcgtgcagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 675
|||||
DB 3047 ACGGCGGTACT-CTCACGGCGCAGCGCATACAGCGCGCGCGCGCGCGCGCGCGAACGCT 2989

QY 676 gccaggttcaactctaccgcgcgcgcgtcgccctccccgcgcgcgcgcgcgcgcgcgcgcgcgc 735
|||||
DB 2988 GGGTGTCTTTCGGCTCCGGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2929

QY 736 gagca 740
|||||
DB 2928 GGCCA 2924

RESULT 8
T61454 standard; DNA; 2539 BP.
ID T61454
AC T61454;
DT 06-OCT-1997 (first entry)
DE Streptomyces viridoporus dhpA gene.
KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW derivative; Streptomyces viridoporus; ester; chiral; synthesis;
KW cardiovascular; treatment; hypertension; ischaemic heart disease;
OS Streptomyces viridoporus.
FH Key Location/Qualifiers

```

FT      338. .2539
FT      /*tag= a
FT      /note="no stop codon given"
FT      950. .2509
FT      /*tag= b
FT      /note="encodes W13666"
PN      WO9705243-A1.
PD      13-FEB-1997.
PD      30-JUL-1996; J02147.
PR      31-JUL-1995; JP-212975.
PR      29-FEB-1996; JP-067478.
PA      (SACC ) MERCIAN CORP.
PA      Aritaawa A, Dobashi K, Isshiki K, Matsufuji M, Nakashima T;
PI      Tsuruta T, Yoshioka T;
PI      WPI; 97-145682/13.
DR      P-PSDB; W13666-67.
DR      Asymmetric hydrolase gene derived from Streptomyces viridosporus -
PT      acts on 4-substituted-1,4-di:hydro:pyridine derivatives to produce
PT      chiral derivatives useful for synthesis of cardiovascular drugs
PT      Claim 3; Page 49-55; 78pp; Japanese.
PT      This sequence is the Streptomyces viridosporus dhpA gene which encodes an
CC      asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
CC      derivatives. The enzyme allows the efficient conversion of 4-substituted-
CC      1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
CC      for use in the synthesis of cardiovascular drugs suitable for the
CC      treatment of e.g. hypertension and ischaemic heart disease.
CC      Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T.
SQ

```

	Query Match	3.9%; Score 75.4; DB 1; Length 2539;	
	Best Local Similarity	42.3%; Pred. No. 8.1e-06;	
	Matches 412; Conservative	0; Mismatches 561; Indels 0; Gaps 0;	
QY	78	ggccacgctacgcagagcgacaggatcgcgcggaacggcccttcggtgcccgcagcgccg 137	
Db	1456	GGACGCGCCCTCACGTCGSGCGCTCGACGACAAGCAAGTCGCGCACTTCTCCTC 1515	
QY	138	gaacgcagaagacaacggcccatggcgttgctgcgcgcgtgaacggcaagacaacgccca 197	
Db	1516	CACCGGCCCCCGCTCGGCAGCGCGCATCAACGCCGACGTCAACGTTCCGGCGTGGA 1575	
QY	198	tggcgtgatgcgcagcgcaacggcaagagacaacggccattggcttgctgcgcagcgaa 257	
Db	1576	CATCACGCGCCCTCGCGGAGGGCAACACATCGGCACGAGGTCGGTGAGGACCAGCGC 1635	
QY	258	cgcacaagacaacggccatccgaggccactgcgaacggccacggccacggccactgcgaa 317	
Db	1636	CGGTACTACATCATCTCCGGCACGTGATGGCCACGCCGACCCCGACGTCGGGGCGGGCGCGC 1695	
QY	318	cggcaagaccacaacggccacgcgcgagagacaacggccattgtagggccgcgcgcgcgaacgg 377	
Db	1696	CCTCTTAAGCAGCAGCACCCCGACTGGACCTCCGCGCAACTGAAGGCGCGCTCAACCGG 1755	
QY	378	cgcagacaacgcagcatgccgaggactccgcgcgcgaacggcgagacaacggcatgcgcgc 437	
Db	1756	CTCCACCAAGGGCGCAAGTACACCCGTTTCGACAGGGTTTCGGSCCGATCCAGGCCGA 1815	
QY	438	ggcggcgacagagagagagggcgttggagtgaattcgcgggtgccaaaggacggcgt 497	
Db	1816	CAAAGCGCTCCAGCAGACCGTGTATCGCGCAGCCCGGTCTCGTGTAGCTTCGGCGTCCAGCA 1875	
QY	498	gctgcggcgacggggcgcaacatgacatccggcgccatacggtaacaagatcacgcgcag 557	
Db	1876	GTGSCCGCACACCGACGACGAGCCGGTCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935	
QY	558	cgtgcaggaagaagggcgcgccccgttgtctgcctggccacggggaccacgctcogttgt 617	
Db	1936	CCAGGACGTACGCTGAAGCTTGACGTGCACCGCCACCCACCCCAAGGGCAAGCGCGCCCC 1995	
QY	618	cccgcccttcgcacgcgcgtcgaggccgagacgcgcgtccgcgcgcgtgtgcacacgg 677	
Db	1996	GGCGGGCTTCTACGCTGGCGCCACACACGGTAGCCCTCCGGCGGGCGGACGCGCTC 2055	

[illegible]

QY	678	ccagttcaactgctacccccgcggcgctgcgctcccgcacgaagcgcgttggcaga	737
Db	2056	CGTCGACATGACCGGCACACCCGGCTCGCGGCCACGGTGGACGGCGACTCGCGGTA	2115
QY	738	geacctgtcgcaggcggtgcgtacatgctgctatcgcgcgaacagcttcttcaccgccg	797
Db	2116	CGTGCTGCCACGGGGCGGGCACACGGTCGCAAGGCCGCCGGGTGCAGCGCAGGT	2175
QY	798	cgggacccaggcgatcgagtgtcataatcccggtgctggcccgagaccgcggcgccaacat	857
Db	2176	CGAGTCGTACGACGTACCGTCCGGGCACATCGCGCGGCACGGCAGCCACGACCGAACA	2235
QY	858	tctgtctcccagggccaggctacccaaactaagcggcgccgctgtacaaggctyga	917
Db	2236	CCTCACCGACCTGATCGGTAGCGGGCTCGGGCTCCGGCGCGGTTACGGCGCCGGC	2295
QY	918	gftccggcatttgacctcatcccgcacaagggtggagatgcatactgcactgcgtgga	977
Db	2296	CACCGACACGGCCACCCTGGCGCTGCCCAAGGGCACCTACCTGTGTGACTCCTGGATCGC	2355
QY	978	atccatcgcgcgaagaacaccaccgcgatgttcataaaccccccaaccgcctgcgg	1037
Db	2356	CANGGACTCGGACGCTCAAGGGCGCATCGACTGGCTGGTCCAGCGCGAGCTGACCGT	2415
QY	1038	cagcgtttactcc	1050
Db	2416	CACCAAGGACACC	2428

RESULT 9

```

RESULT          9
T61455
ID ID T61455 standard; DNA; 2809 BP.
T61455;
AC AC 06-OCT-1997 (first entry)
DT DT
DE DE DhpA-mel chimeric gene.
KW KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
QS QS Chimeric Streptomyces viridosporus.
OS OS Chimeric Streptomyces antibioticus.
FH FH Key Location/Qualifiers
FT FT cds 338..2809
FT FT /*tag= a
FT FT 338..2539
FT FT /*tag= b
FT FT /product= DhpA_protein_product
FT FT /note= "from S. viridosporus"
FT FT 2540..2809
FT FT /*tag= c
FT FT /product= Melanin
FT FT /note= "from S. antibioticus"
PN PN W09705243-A1.
PD PD 13-FEB-1997.
PF PF 30-JUL-1996; J02147.
PR PR 31-JUL-1995; JP-212975.
PR PR 29-FEB-1996; JP-067478.
PA PA (SAOC ) MERCIAN CORP.
PI PI Arisawa A, Dobashi K, Isshiki K, Matsufuji M, Nakashima T;
PI Tsuruta T, Yoshioka T;
DR DR WPI: 97-145682/13.
DR DR P-PSDB; W13668.
PT PT Asymmetric hydrolase gene derived from Streptomyces viridosporus -
PT acts on 4-substituted-1,4-di:hydro:pyridine derivatives to produce
PT chiral derivatives useful for synthesis of cardiovascular drugs
PT Claim 5; Page 37-43; 78pp; Japanese.
PS PS This sequence is a fusion gene encoding Streptomyces viridosporus dhpA
CC CC gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-
CC CC dihydropyridine derivatives, and melanin from S. antibioticus. The DhpA
CC CC enzyme allows the efficient conversion of 4-substituted-1,4-
CC CC dihydropyridine esters to chiral partially hydrolysed derivatives,
CC CC for use in the synthesis of cardiovascular drugs suitable for the
CC CC treatment of e.g. hypertension and ischaemic heart disease.
SQ SQ Sequence 2809 BP; 473 A; 1041 C; 984 G; 311T;

```

[illegible]

```

RESULT 10
V05372 standard; RNA; 2277 BP.
ID V05372 standard; RNA; 2277 BP.
AC V05372;
DE 06-JUL-1998 (first entry)
DE Human telomerase p105 subunit mammalian optimised synthetic RNA.
KW telomerase; p105; human; cell replication; cancer; restenosis;
KW multiple sclerosis; inflammation; rheumatoid arthritis;
KW myocardial infarction; glomerulonephritis; transplant rejection;
KW infection; therapy; ds.
KW Homo sapiens.
OS Synthetic.
PN WC9801542.A1.
PD 15-JAN-1998.
PF 08-JUL-1997; U12296..
PR 08-JUL-1996; US-676974.
PA (REGC ) UNIV CALIFORNIA.
PI Collins K;
DR WPI; 98-101043/O9.
PT New nucleic acid encoding human telomerase proteins or their
PT fragments - useful for therapeutic modulation of telomerase activity
PT and for screening for potential modulators of telomerase-target
PT binding
PS Disclosure; Page 23-24; 32pp; English.
CC This polynucleotide comprises a synthetic RNA sequence coding for
CC the p105 subunit (see W46593) of human telomerase and optimised for
CC expression in mammalian cells. It is based on an isolated cDNA
CC clone (see V05369) for p105. The invention provides methods relating
CC to human telomerase and related nucleic acids, including the subunit
CC proteins p140, p105, p48 and p43. The proteins may be produced
CC recombinantly from transformed host cells or purified from human
CC cells. Also included are human telomerase RNA (see V05373) and
CC functional derivatives (see V05374 and V16092-93), as well as p105
CC synthetic DNA sequences (V05370-72). The invention also provides
CC isolated telomerase hybridisation probes and primers capable of
CC specifically hybridising with the telomerase gene, telomerase-
CC specific binding agents such as specific antibodies, and methods of
CC making and using the subject compositions in diagnosis (e.g.
CC genetic hybridisation screens for telomerase transcripts), therapy
CC (e.g. gene therapy to modulate telomerase gene expression) and in
CC the biopharmaceutical industry (e.g. reagents for screening
CC chemical libraries for lead agents). Modulation of telomerase
CC expression can be used for the treatment or prevention of cancer,
CC restenosis, inflammation, myocardial infarction, glomerulonephritis,
CC transplant rejection or infections (e.g. with HIV). 216 T;
SQ Sequence 2277 BP; 557 A; 723 C; 781 G; 216 T;

Query Match          3.9%; Score 74.2; DB 1; Length 2277;
Best Local Similarity 45.8%; Pred. No. 1.3e-05;
Matches 333; Conservative      0; Mismatches 388; Indels    6; Gaps     2;

QY   75 catggccaccgtacgcagacgaggcagtgctcgcgaaacggccttgcgtggccgcagc 134
       ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db   585 CAAGGACCACCGAGGTGGTGCATCGGCCGAGGAAAGAACCAAGACGACGAACCA 644
       ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY   135 cgcgaaacgaagaagcaaacgccatggctggctgcccgcgtgaacggcaagacaacgg 194
       ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db   645 GGAGAGCGTGTAAGAAGAAGGGCCCGAGGAGGAGGACATGGAGGAGGAGAGACGACA 704
       ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY   195 ccattggcgtgatgccacgcgcaacggccaagagacaacggcccatggctggtcgccgaogc 254
       ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db   705 CGACGACGAGCACGACACGAGGAGGACGGCGCTGTTTCGACACGACGAGGACGAGGAGGA 764
       ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY   255 gaacgcgaacgaacaaacgg----ccatgccaggccacctgcgaacggccccgcaggccac 311
       |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db   765 GAACATCGAGGACCAAGGTGACCAAGCCCGTGCATCTCCAGNAGGCGCCGTGTGAGCGCCC 824
       |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY   312 tcgcaacgcgcaagaccacacggccacgcgagagacaacggcccatgctgtaggcgcgcagc 371
       |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db   825 CGCCCCCGCCCAAGAGCAGCAGCACACACGAGGAGGACGAGGACGAGCCTTGGAGGAGCGCAG 884
       |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

```

```

Query Match      3.8%; Score 73; DB 1; Length 114955;
Best Local Similarity 30.8%; Pred. No. 4.2e-05;
Matches 264; Conservative 97; Mismatches 494; Indels 3; Gaps 2

QY 67 cgttctcatggtccaccctagccagagcaacgagtcgcgcgaacggccttgccgtg 126
   || || || || || || || || || || || || || || || || || || || || ||
Db 104880 CGGNHHNNNSVGGCCVCGCGGNHHNNNSVGGCCVCGCGGNHHNNNSCGVCGCCVCGGNNH 104821

QY 127 gccgcagcgcgcaagcgcaagcgcaacggccatgctggtgcgcgcgtgtaacggccaag 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104820 NNNSCCVCGGCCVCGCGGNHHNNNSGCCVGGCCVCGCGGNHHNNNSCGCCVCGGCVCVCGG 104761

QY 187 agcaacggcgcgtggtgatccgcacgcgaacggcaagagcaacgcccatggcgtgacct 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104760 NNHHNNSGCCGCCVCGCGGNHHNNNSCGCCGCCVCGCGGNHHNNNSGCGCGC 104701

QY 247 gccgcagcgaacggcaagcgcaacggccatgccagaggccactgcgaacggccacggcgag 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104700 CVCVGGCVCVCGGNHHNNNSGGCGCCVCGGCCVCGCGGNHHNNNSCGCGCCCGCVCVGGCC 104641

```

QY 307 gcaatgcgaacggcaagaccacgccaacccgccagcagaacaaggccatgtctagggccgc 366
Db 104640 VGGGNNHNHNSGCCGCGCCCGVCGCGGNNHNHNSGGCGCGCCGCVCVCVGC 104581

QY 367 gcgcgaacggcgagagaaacgagcatgcgagagactccgcggcgcaacgcgcgagagcaac 426
Db 104580 GGNHNHNSGCCGGCGCGCCCVGGCCVCGCGNHNHNSCCGCGCGCGCCCVGGCCVG 104521

QY 427 ggg--catgcggcgggcg -cgagagagagagagcggtgagtggaaatttcgcgggt 483
Db 104520 CGNHNHNSGCCGGCGCGCCCGVGGCCVCGGNHNHNSGGCGCGCGCGCCVG 104461

QY 484 gccaaagcacggctgtctgcccgcacggggcgcaaacatgagcatccggcgcatcacggtac 543
Db 104460 GCVCGCGNHNHNSGCCGCGCGCGCGCCVGGCCVCGGNHNHNSCCGCGCGCGG 104401

QY 544 aagatcagcgcgagctgcagagagaaggcgccgcgcgctgtgcgctggccacagg 603
Db 104400 CGCGCCVGGCCVCGGNHNHNSCCCGCGCGCGCGCCVGGCCVCGGNHNHNS 104341

QY 604 gaccctccgttgttccoggccttcgcacggccgctgagcgagcgagacocgtgcgcgc 663
Db 104340 GCCGCGCGCGCGCGCCVGGCCVCGGNHNHNSGGCGCGCGCGCGCGCGCCVG 104281

QY 664 gccgtgcgcaccggccagttcaactctaccccgcgcgctccgcctccccgcgcacga 723
Db 104280 GCCVCGGNHNHNSGGCGCGCGCGCGCGCGCCVGGCCVCGGNHNHNSGCCGG 104221

QY 724 agcgctgtgcagagacctgtgcagggtgcgttacatgtatcgcgcacgacgtc 783
Db 104220 CCGGNHNHNSGCCGCGCCGANHNHNSGCCCGCGCGACNHNHNSGCCGCGCCGAC 104161

QY 784 ttctcaccgcggcggggaaccaaggcgatcgaggtcataatccgggtgctgcccagacc 843
Db 104160 GNHNHNSGCCCGCGCGAGCGCNHNHNSGCCGCGCGCGCGACGCNHNHNSGCCGCGCC 104101

QY 844 gccggcgccaacattgtctcccgacggccagggtaccaaaactacgagcgcgcgcgcg 903
Db 104100 GACGCGGNHNHNSGCCGCGCGCGACCGCGCNHNHNSGCCGCGCGCGCGCCNNH 104041

QY 904 ttcaacagggttgaggtc 921
Db 104040 NNSGCCGCGCGACGCC 104023

RESULT 13

X53491
ID X53491 standard; DNA; 114955 BP.

AC	X53491;
AD	05-JUL-1999 (first entry)
DE	Human adenosine A1 receptor antisense oligonucleotide fragment.
KW	Antisense oligonucleotide; multiple target; antisense treatment;
KW	impaired respiration; inflammation; lung disease;
KW	pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW	acute asthma; allergy; asthma; impaired respiration;
KW	respiratory distress syndrome; pain; cystic fibrosis;
KW	pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW	colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW	prostate cancer; ss.
OS	Synthetic.
PN	W09913886-A1.
PD	25-MAR-1999.
PF	17-SEP-1998; U19419.
PR	09-JUN-1998; US-093972.
PR	17-SEP-1997; US-059160.
PA	(UYEC) UNIV EAST CAROLINA.
PI	Nyce JW;
DR	WPI; 99-229400/19.
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT	vasoconstriction
PS	Disclosure; Page 37; 120pp; English.

The specification describes antisense oligonucleotides (X52869-X55271) directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences X5272-74. These multiple target oligonucleotides (specifically X55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis, acute asthma, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer. Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; SQ

Query Match 3.68; Score 69.6; DB 1; Length 114955;

Best Local Similarity 35.0%; Pred. No. 0.00018;

Best local similarity	Seq. no.	Seq. no.	Seq. no.
Matches 230; Conservative 44; Mismatches 379; Indels 5; Gaps 1;			

Qy 95 gcgacggagtgcgccgcgaacggccttgccgtggccgcagccgcgaacggcaagcaacg 154

D_b 108402 GCGCGCCTGGCTCGCSNNNDNNGCGGGCGCGCTGGCTCGSNNNDNNGCGG 108461

Qv 155 gccatggcgtggctgcccgcgtgaacggcaaggcaacggccatggcgtggatgccgacg 214

Db 108462 GGGCGCGCGCGCTCSNNNDNNGCGCGCGCGCTGGCTSNNDNNGCG 108521

(v) $\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2} dx = f(0)$

108587

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1990	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100											

LD I0804Z C6CCSNNNDNNGGCGGGGGCGGCSCSNNNDNNGGCGGGGGCGGCSCSNNNDNNGG 100700Z

Qy

395 ccgaggactccgcggcggaacgagcagagcaacgggccatgcggcgccgagcaggaggg
| | | | : | | | : | | | : | | | :

DB T08/02 CGGGGGCGCGCSNNNDNNGCGGGGGCGCGGSNNNDNNGCGGGGGCGCGGSNNNDN T08/01

Qy 455 aggaggcggaggaggaaatttcgcgggtgcgaaggacggcgtgctggcgacggggg 514

Db 108762 NGCGGGGGCGGCSNNNDNNGCGGGCGGGSNNNDNNGCGGGCGGCTGGCT 108821

QY 515 cgaacatgagcatccgggcgatacgggtacaagatcagcgagcgtgcaggagaaggggc 574

Db 108822 CGCCTBGGGGCCCCSNNNDNNCGGGGGCGCGCCTGGCTCGCCTBGGGGCCCCSNNNDN 108881

Qy 575 cgcggccgtgtgcgcgtggccacgggacccgtccgtgttcccg ----gccttccg 629

Db 108882 NGGGGGCGGCGCCTGGCTCGCCTBGGGCCCSNNNDNNGGGCGGCGCCTGGCT 108941

Qy 630 cacqccgtcgaqqccgaqqacgccctcgccqccgcactgcaactg 689

Db 108942 CGCCTBGGGCCCCSNNNDNNGGGGGGGGGCGCTGGCTCGCCTBGGGCCCCSNNNDNNGG 109001

690 ctaccgcgcgcattcggcctccccccgcacgaagcgcctgacagagacacctatca 747

DB 109003 CCCCCCTCCTCCCTCCCCCNNDNNCCCCCTCCTCCTCGCTBC 109059

RESULT 14

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 15:03:30 ; Search time 1096.08 Seconds
(without alignments)
3437.285 Million cell updates/sec

Title: US-09-026-400-3
Perfect score: 1910
Sequence: 1 cgcgctactagtagtattcc.....aagttaaaaaaaaaaaaa 1910

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*

1: em_est1:.*
2: em_est2:.*
3: em_est3:.*
4: em_est4:.*
5: em_est5:.*
6: em_est6:.*
7: em_est7:.*
8: em_est8:.*
9: em_est9:.*
10: em_est10:.*
11: em_est11:.*
12: em_est12:.*
13: em_est13:.*
14: em_est14:.*
15: em_est15:.*
16: em_est16:.*
17: em_est17:.*
18: em_est18:.*
19: em_est19:.*
20: gb_est1:.*
21: gb_est2:.*
22: gb_est3:.*
23: gb_est4:.*
24: gb_est5:.*
25: gb_est6:.*
26: gb_est7:.*
27: gb_est8:.*
28: gb_est9:.*
29: gb_est10:.*
30: gb_est11:.*
31: gb_est12:.*
32: gb_est13:.*
33: gb_est14:.*
34: gb_est15:.*
35: gb_est16:.*
36: gb_est17:.*
37: gb_est18:.*
38: gb_est19:.*
39: gb_est20:.*
40: gb_est21:.*
41: gb_est22:.*
42: gb_est23:.*
43: gb_est24:.*
44: gb_est25:.*
45: gb_est26:.*
46: gb_est27:.*
47: gb_est28:.*
48: gb_est29:.*
49: gb_est30:.*
50: gb_est31:.*
51: gb_est32:.*
52: em_est20:.*
53: em_est21:.*

54: em_est22:.*
55: em_est23:.*
56: em_est24:.*
57: em_est25:.*
58: em_est26:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	372.6	19.5	571	49	AI657244	AI657244 486092E01
2	313.8	16.4	594	49	AI619119	AI619119 486073F11
3	256.2	13.4	459	21	D48521	D48521 RICS14756A
4	233.8	12.2	410	20	D25143	D25143 RICS343A R
5	182.4	9.5	464	49	AU056779	AU056779 AU056779
6	151.6	7.9	446	20	Z38084	Z38084 ATTS422A Ve
7	132	6.9	470	23	T76370	T76370 11148 Lambd
8	121	6.3	352	23	T13684	T13684 1849 Lambda
9	118.4	6.2	343	20	D23043	D23043 RICC2086A R
10	108.8	5.7	283	49	AI621525	AI621525 486092E01
11	102.2	5.4	719	40	AA980369	AA980369 ua52c12.r
12	92.4	4.8	512	34	AA511748	AA511748 v129f11.r
13	89.8	4.7	405	20	Z38047	Z38047 ATT34207 Ve
14	85.2	4.5	415	48	AI597379	AI597379 v129f11.y
15	80.4	4.2	347	23	T41757	T41757 10338 Lambd
16	78.2	4.1	801	42	AI116859	AI116859 ue29d01.y
17	77	4.0	559	33	AA394723	AA394723 26506 Lam
18	75.6	4.0	729	41	AI048570	AI048570 ud61e01.y
19	75	3.9	757	40	C91185	C91185 C91185 Dict
20	74.4	3.9	536	47	AI487927	AI487927 EST146249
21	69.8	3.7	664	41	AI055475	AI055475 coau0004B
22	66.6	3.5	870	48	AI562691	AI562691 TENS2678
23	64.6	3.4	779	48	AI562604	AI562604 TENS2761
24	61.8	3.2	521	42	AI116355	AI116355 uf02b01.y
25	61.6	3.2	437	20	T53940	T53940 yb85b03.r1
26	61.6	3.2	549	26	W43288	W43288 22661 Lambd
27	61.4	3.2	699	45	AI386363	AI386363 mo05a03.y
28	57.6	3.0	528	25	N97159	N97159 22338 Lambd
29	57.2	3.0	820	41	AI001336	AI001336 MEST6-A6.
30	56.2	2.9	288	36	AA650709	AA650709 30823 Lam
31	55.4	2.9	470	37	AB009149	AB009149 AB009149
32	54.8	2.9	248	36	C72041	C72041 C72041 Rice
33	54.8	2.9	641	45	AI357868	AI357868 qv13b02.x
34	54.2	2.8	676	45	AI389106	AI389106 GH20192.5
35	53.8	2.8	581	41	AI063350	AI063350 GH03102.3
36	53.4	2.8	806	47	AI523757	AI523757 tg94f08.x
37	53.4	2.8	610	48	AI609782	AI609782 t183a03.x
38	53.4	2.8	598	49	AI658863	AI658863 tt98f04.x
39	53	2.8	833	43	AI189624	AI189624 qd32h08.x
40	53	2.8	640	49	AU056191	AU056191 AU056191
41	52.8	2.8	503	43	AI184898	AI184898 oo07h05.x
42	52.8	2.8	539	43	AI192517	AI192517 qe61c01.x
43	52.8	2.8	449	45	AI369183	AI369183 q775b04.x
44	52.8	2.8	599	46	AI453638	AI453638 tj45g10.x
45	52.8	2.8	448	50	AI694026	AI694026 wd13h09.x

ALIGNMENTS

RESULT 1

AI657244

LOCUS

DEFINITION

ACCESION

NID

VERSION

AI657244 571 bp mRNA
486092E01.y1 486 - leaf primordia cDNA library from Hake lab Zea
mays cDNA, mRNA sequence.
AI657244
94753339
AI657244.1 GI:4753339

```

KEYWORDS      EST.
SOURCE         Zea mays.
ORGANISM       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
                Poaceae; Zea.
REFERENCE      1 (bases 1 to 571)
AUTHORS        Walbot,V.
TITLE          Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL        Unpublished (1999)
COMMENT        On Jun 5, 1998 this sequence version replaced gi:3189324.

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 486092 row: E column: 01.
Location/Qualifiers
1. 571
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="486 - leaf primordia cDNA library from Hake
lab"
/tissue_type="leaf primordia"
/dev_stage="P7-P11 leaf"
/lab_host="E.coli XL1-Blue MFR"
/notes="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
library."

BASE COUNT    108 a 198 c 178 g 87 t
ORIGIN

Query Match    19.5%; Score 372.6; DB 49; Length 571;
Best Local Similarity 80.3%; Pred. No. 1e-53;
Matches 462; Conservative 0; Mismatches 109; Indels 4; Gaps 2;

QY 541 tacaatcagcgcagcgtgcaggagaagggccggcccgccgtgctgcgctggccac 600
Db 1 TTCAAGATCAGCCGAGCCTTGGACGCGCGGACCCGCGCCCGCTGCTGCGCTGGCGCAC 60

QY 601 ggggaccgcgtgtgttcccgcccttcgcacgcccgtgcagccgagccgagcgcgtcgcc 660
Db 61 GGGGACCCCTCGGTGTTCGGCGCTTCCGACCGCGCGGAGCCGAGGCGCGGTGGCC 120

QY 661 gcgcgcgtgcgacccgcccagttcaactgctaccgccggcgctgcgcctcccgcgcga 720
Db 121 GCGCGCTCCGGACCGGCAAGTTCAACTGTACCCGCGCGGCTCGGCTCCCGAAGCC 180

QY 721 cgaagcgcgtgacagcacctgtgcagggcggtgcgtacatgctatcgccacgac 780
Db 181 CGCGGTGCTCTGCGGAGGACCTTGTGAGGACCTTCCATACAGTGTGCGAGCGACGAC 240

QY 781 gtcttctcacgcgcggcgagcccccagggcgatcgaggtcataatcccggtgctggcccag 840
Db 241 ATCTTGTCAACCGCGAGGACGCGAAGCATTGAGGTGGTGTCTCAGTCTCGGCGAG 300

QY 841 accgcggcgccaaattctgtctcccgagcaggtatcccaaatcagagcgcgcc 900
Db 301 ---CCGGGACCAACATCTGTCTCCGAGCGGGGTATCCGAACATACGAGCGCGCA 357

QY 901 ggttcaacaggtgaggtgcggcatttcgacctatcccgacaaaggggtggagatc 960
Db 358 GGGCTCCACAACCTGGAATTTCGCGGTTTCATCTGATCCCCGAGAGAGGGTGGAGATT 417

QY 961 gacatgactgctggaattccatcgccgacaagaacaccaccgcatggtcatcataaac 1020
Db 418 GACATGACGGGTGTGGATCGATCGCCGACGAAGAACACCACCGCCATGTGTCATATAAC 477

```

```

QY 1021 ccacaaccccgcgcgagcgtttactctacgaccattgtccaaggtcgcggaggtg 1080
Db 478 CCACAACCCCTGCGGGAGTGTCTACACCCGAGACATTTGGCCAAAGTTCGGGA-GTG 536

QY 1081 ggaagaagcgtcggaattgtgtgattgctgacga 1115
Db 537 GCAAGGAAGCTTGAATACTGGTCATCGCTGATGA 571

RESULT 2
AL019119/c
LOCUS
DEFINITION    486073F11.x2.486 - leaf primordia cDNA library from Hake lab Zea
                mays cDNA, mRNA sequence.
ACCESSION     AI619119
NID           94628245
VERSION       AI619119.1 GI:4628245
KEYWORDS      EST.
SOURCE        Zea mays.
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
                Poaceae; Zea.
REFERENCE      1 (bases 1 to 594)
AUTHORS        Walbot,V.
TITLE          Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL        Unpublished (1999)
COMMENT        On Mar 10, 1998 this sequence version replaced gi:2948200.

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 486073 row: F column: 11.
Location/Qualifiers
1. 594
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="486 - leaf primordia cDNA library from Hake
lab"
/tissue_type="leaf primordia"
/dev_stage="P7-P11 leaf"
/lab_host="E.coli XL1-Blue MFR"
/notes="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
library."

BASE COUNT    172 a 113 c 112 g 197 t
ORIGIN

Query Match    16.4%; Score 313.8; DB 49; Length 594;
Best Local Similarity 73.8%; Pred. No. 7.4e-44;
Matches 399; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 1276 caggaaactaagatctctacatcaattcgaattaccctcaatgtctgcagacccagca 1335
Db 593 CAAGAAACCAAGATCATTCGATCGATCACAACACTTCCCTTAACGTTTCAACAGATCCGGCA 534

QY 1336 accttcattcggcagctctctcagattcttgagaacacaaagaagattctttaaag 1395
Db 533 ACTTTTGTTCAGGAGCGCTTCCCATATTCTTAGACACACAAAGAGATTTTTCAG 474

QY 1396 gcatattgtctgtctaaaggaatcatcagagatatgtctacaaataaaggaaac 1455
Db 473 AGGATCATTTGTTCTGCTAGCGGAAACATCAGAGATATGTTTCAGCGGAATAAAGGACATC 414

QY 1456 aaatcattacatgctccacaagccagaagatcaatgtttgtcatgtgtaaacgac 1515
Db 413 AAGTGCATCATATGCCCTCCCAAGCCAGAGTTCCAAGTTTGTGTCATGTTGTAATAACTAAAT 354

```

```

QY 1516 ttacatcttttggagaaatagacatgacattgatttctgctgaagctcgaataaa 1575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 TTGTATCTTTGGAGAGCATCCATGATGATATGATTTTGTGAAGTGGCAAGAA 294

QY 1576 gaatcagtaattttatgccccggaggtgtcttggaaatggcaaaactgggtccgcattact 1635
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 GAGTCGGTGATTTGTGTCCAGGGAGTGTGTTGGGAATGGAAACTGGATCGGTACAT 234

QY 1636 ttgtctgtgtccatctctctcaagatgctctcggaagatcaaatcattctgtcaa 1695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 TTCGCCATTGATTCATCTCTCTCTGTGATGGTCTGTGAGAGGCTGAAATCTTTCGCCAA 174

QY 1696 aggaacaagaagaataatctgcagcagatgattgttagtttatctgactgaagctgttaa 1755
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 AGGCATAAGAAGAAGATTGCTTAATGCCATTACTATATTCGACTTCAAAAGTTGTTA 114

QY 1756 atcattccagtatccccatctatatcttctcaataaaatggaaactttttagttctctatga 1815
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 CCCACTTCCATCGTTTTCCTCCCTGTATCATCAACAAAGTGTAACTTGAATCCCTCAGA 54

QY 1816 a 1816
Db 53 A 53

RESULT 3
D48521 48521 459 bp mRNA EST 02-AUG-1995
LOCUS R1CS14756A Rice green shoot Oryza sativa cDNA, mRNA sequence.
DEFINITION D48521
ACCESSION D48521
NID 9702230
VERSION D48521.1 GI:702230
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 459)
AUTHORS Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE Rice cDNA from callus 1995
JOURNAL Unpublished (1995)
COMMENT

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
Insert Length: 852 Std Error: 0.00
High quality sequence stop: 405.

FEATURES
    source
        1..459
        /organism="Oryza sativa"
        /strain="Nipponbare"
        /db_xref="taxon:4530"
        /clone_lib="Rice green shoot"
        /note="Green shoot (8 days old)"
BASE COUNT 134 a 86 c 105 g 132 t 2 others
ORIGIN

Query Match 13.4%; Score 256.2; DB 21; Length 459;
Best Local Similarity 73.5%; Pred. No. 3.3e-34;
Matches 339; Conservative 0; Mismatches 120; Indels 2; Gaps 1;

QY 956 agatcgactcgactcgctggatccatcgacgaagaacaccaccgcctggtgcatca 1015
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 AGATTGACCTTACTCCCTAGATCTATTGCGGACGAAGACACTACTGCGGATAGTCATCA 60

```

```

QY 1016 taaccccccaaacccgtgcgcagctttactcctacgaccatctgtccaagctcgcg 1075
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TAAATCCCAATAATTCATCGGGGAATGTGTACACTTACGAGCATTTATCCAAGGTGGCAG 120

QY 1076 aggtggcgaaaaaggttcggaatattgttgatgctggaagaggtatatacggaagctggctc 1135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 AGGTAGCAAGGAAGCTTGGGNTATTGGTAATTACTGATGAGGTGATGTAATTTGGTTT 180

QY 1136 tgggcagggccccgttccatcccaatgggaggtgttggggcaaatcaccctgtgctgtcca 1195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 TTGGGAGTTCCCAATTTGTCCCAATGGGTTGCTTTGGGCACATCGTACCAAATATTAAACA 240

QY 1196 taggtctctctgtcccaagtcattgtagtcctggatggcggttggatgggttagcggtgt 1255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TAGGATCGCTATCAAGAGGTGGATAGTCCCGGGATGGGACTTGGTGGGTAGCAATAT 300

QY 1256 acgaccccgaaagatctttacaggaataactaatgattctctacatacaattacgaattaccta 1315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GTGACCCCAAGAAGACTCTACAAGAAACCAAGNTTGCAACATTAATTACTAATTTCTCTTA 360

QY 1316 atgtctcgacagaccagcaaccttcattcaggcagctctctcagattcttgagaaca 1375
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 ATGTTTCAACTGATCCGCAACTTTTCATTTCAGGGAGCTCTACCGAATATTTCTTAAGATAC 420

QY 1376 caaaggaaagattctttaaggcgattattgtctgctgctaaag 1416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 CAA--GGAAGATCTTTAAGGGGTAATTGATTTCGTTACG 459

RESULT 4
D25143 48521 410 bp mRNA EST 20-JUL-1998
LOCUS R1CR3343A Rice root Oryza sativa cDNA clone R, mRNA sequence.
DEFINITION D25143
ACCESSION D25143
NID 9428987
VERSION D25143.1 GI:428987
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 410)
AUTHORS Minobe,Y. and Sasaki,T.
TITLE Rice cDNA from root
JOURNAL Unpublished (1995)
COMMENT

Contact: Yuzo Minobe
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: minobe@rtes0.riken.go.jp
PROJECT = "RGP".

FEATURES
    Location/Qualifiers
        1..410
        /organism="Oryza sativa"
        /strain="Nipponbare, sub_species Japonica"
        /db_xref="taxon:4530"
        /clone_lib="R"
        /clone_lib="Rice root"
        /note="Prepared from seedling root."
BASE COUNT 125 a 69 c 97 g 119 t
ORIGIN

Query Match 12.2%; Score 233.8; DB 20; Length 410;
Best Local Similarity 74.8%; Pred. No. 1.9e-30;
Matches 306; Conservative 0; Mismatches 102; Indels 1; Gaps 1;

```

QY 1072 gcggaggtgcccgaagctcgcgaataattgttgatgctgaacaggtatatacggcaagctg 1131
 Db 2 GCAGAGGTACCAAGGAAGCTTGGATATTGGTAATTACTGATGAGGTGTATGTAATTG 61
 QY 1132 gttctgggcagcgcgcgcgttccatcccaatgggagtggttttgggcacatcacccctgtgctg 1191
 Db 62 GTTTTGGGAGTTCCCATTTGTCCTCCCAATGGTGTCTTTGGGCACATCGTACCAATATTA 121
 QY 1192 tccatagggtctgttccaaagtcatgtagtagtgcctggatggcggcgttgatgggtagcg 1251
 Db 122 ACCATAGGATCGCTATCAAAAGAGTGGATAGTGCCTGGGATGGGACTTGGTTGGTAGCA 181
 QY 1252 gtgtacagccccgaagaagatcttcacaggaactaaagatctctacatcaataacgaattac 1311
 Db 182 ATATGTGACCCCAAGAAAGACTCTACAGAAACCAAGATTGCCACATTAATTACTAATTC 241
 QY 1312 ctaatgtctgcagacaccagcaacaccttcattcaggcagctcttctcagattcttgag 1371
 Db 242 CTTAATGTTTCAACTGATCCAGCAACTTTCATTACGGGAGCTCTACCGAATATCTTAAG 301
 QY 1372 aacacaaagaagattctttaaggcattattggtctgc-taaaggaatcatcagagat 1430
 Db 302 AATACCAAGGAAGAAATCTTTTAGAGGATAAATTGATTGCTTACGGGAAACATCAGATAT 361
 QY 1431 atgctacaaagaataaagaaacaaatacatattacatgctcctcacaag 1479
 Db 362 TTGCTATAGAGGATAAAGGTTAATAATGCATCACTTGTCCTCACAAAG 410
 RESULT 5
 LOCUS AU056779 464 bp mRNA EST 29-APR-1999
 DEFINITION AU056779 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA clone S20870_1A, mRNA sequence.
 ACCESSION AU056779
 NID 94715663
 VERSION AU056779.1 GI:4715663
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
 1 (bases 1 to 464)
 YAMAMOTO, K. and SASAKI, T.
 Rice cDNA from mature leaf
 Unpublished (1999)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3187002.
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = 'RGP'.

FEATURES source

Location/Qualifiers
 1..464
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /map="21q"
 /clone="S20870_1A"
 /clone_lib="Oryza sativa mature leaf Nipponbare"
 /tissue_type="mature leaf"
 BASE COUNT 117 a 109 c 105 g 132 t 1 others
 ORIGIN

Query Match 9.5%; Score 182.4; DB 49; Length 464;
 Best Local Similarity 63.4%; Pred. No. 8.1e-22;
 Matches 295; Conservative 0; Mismatches 167; Indels 3; Gaps 1;
 QY 687 ctgctaccctccgcgcgtcgctggcctcccccgcgcaagaagcgcgtggcagacacctgtc 746
 Db 1 CGGCTACNCTCCCGCCGCCACGAGCTCGCTGCTGCCGAGCTATTGACAGATACCTATC 60
 QY 747 gcaggcgctgcgtacatctatcgcgcgaacgctcttcctcacgcgcgcgggagcca 806
 Db 61 CTGTGATCTTCCTTACAAGCTTGCACAGATGATATTTTCTCACATCTGGAGGTACCCA 120
 QY 807 ggcgatcgcaggtcctaataatccccgtgctggccagacgcgcgcgccaacatctctctccc 866
 Db 121 AGCAATCGAGATTGTATGCTGTTTTGGCCA---ACCAGGTGCCAATATATTGCTTCC 177
 QY 867 caggccagctaccacaaactacagggcgcgccgcggttcaacagcgtggaggtccggca 926
 Db 178 AAAGCCGGGTACCCAAAACATGAACACATGCGGTGTTTCCACAGGATGGAAGTGGGCT 237
 QY 927 ttctacatctcccgcacaagggatgggagatcacatcgactcgctggaatccatcgc 986
 Db 238 CTATGATCTGTTCAGAGAGAGATGGAGATTATGTTGAAGCTGTTGAAGCTTTAGC 297
 QY 987 cgacaagaacacccaccgcctatggtcatcataaaccacaccccggtggcgaggttta 1046
 Db 298 AGATGAGAATACTGTGCAATAGTATTACTAACCCCAATAACCCCTTGTGGTAATGTGA 357
 QY 1047 ctctacgacctctgtccaaagtcgcgaggtggcgaaagcgttcggaattatgtgat 1106
 Db 358 CACTTATGAGCATCTGTCCAAAGATTGCAGATACAGCAAGATTCGGTCTCTGTAGTCAT 417
 QY 1107 tgcctacgaggtatcgcgaactggtctctggcagcgcgccgtt 1151
 Db 418 TGTGATGAATATATGTCACCTTGTATTATGAAGCACTCCTTT 462
 RESULT 6
 LOCUS Z38084 446 bp mRNA EST 24-OCT-1994
 DEFINITION AT9S4224 Versailles-VB Arabidopsis thaliana cDNA clone VBMA10 5', mRNA sequence.
 ACCESSION Z38084
 NID 9558161
 VERSION Z38084.1 GI:558161
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 446)
 CNRS.
 The Arabidopsis thaliana transcribed genome: the GDR cDNA program
 Unpublished (1996)
 COMMENT
 Contact: Desprez T., Amselem J., Chiapello H., Rouze P., Caboche M., Hofte H.
 INRA Versailles
 Laboratoire de Biologie Cellulaire
 Route de Saint-Cyr, 78026 Versailles Cedex, France
 Email: thierry@versailles.inra.fr.
 Location/Qualifiers
 1..446
 /organism="Arabidopsis thaliana"
 /strain="ecotype Columbia"
 /db_xref="taxon:3702"
 /clone="VBMA10"
 /clone_lib="Versailles-VB"
 /note="Vector: pBluescript; tissue-whole seedlings of A. thaliana ecotype Columbia; clone_library=versailles-vb; Cloning vector: pBluescript (Stratagene); Physiological

conditions: in vitro-grown etiolated seedlings, 5 days old."

BASE COUNT 100 a 81 c 111 g 146 t 8 others
ORIGIN

```

Query Match      7.9%; Score 151.6; DB 20; Length 446;
Best Local Similarity 61.0%; Pred. No. 1.2e-16;
Matches 272; Conservative 0; Mismatches 172; Indels 2; Gaps 2;

QY 913 ctgaggtcggcatttcagctcatcccgacaaaggggtggagatcgacatcgactcg 972
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CTNGAANTTCGCTACGTCGATCTCTTCGCGAAATGGATGGAGATCGATCTTATGNN 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 973 ctggaatccatcccgacaaagacaccccgccatg-gtcatcataaaccccaacaccc 1031
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GTCGAGGCTCTTCAGACGAAACACAGGTTGCTTTTANTGTATTAACCTGGTATCC 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1032 gtgcgcagcgtttatctctacgaccatctgtccaaggtgcggaggtggcgaaggct 1091
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TTCGCGGAATGTCTATAGCTACCAAGCATTTGATGAAGATTGCGGAATCGCGGAAAAA 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1092 cggaaatggtgatctgacgaggtatagcgaagctggttctggcagcgcccggtt 1151
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 AGGTTCTTGATTCGATTCGATGAGGTTACGGTCATCTGCTTTGGTAGCAAAACCGTT 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1152 catcccaatggggagtggtgggcacatcacccctgtgctgtccataggggtctctgtccaa 1211
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TGTGCCAATGGGNGTGTGGATCTATTGCTTNCCTGTGCTACCTTGGCTCTTATCAA 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1212 gtcattgatgtcctggatggtggtggtggtggtggtggtggtggtggtggtggtggt 1271
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GAGATGGATAGTTCCAGGTTGGGACTCGGGTGGTGTGTGTCACCACTGATCCTTCTGTT 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1272 c-ttacagaaactaagactctacatcaattacgaattacatcctcaatgtctgcagagacc 1330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CTTTAAGGACCTTAAGATCATTGAGAGGTTTAAGAAATACTTGTATTTCTTGGTGGAC 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1331 cagcaaccttcattcaggcagctctt 1356
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 CAGCTACATTATTACGGCTGCAGTT 446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
T76370 470 bp mRNA EST 09-JAN-1998
LOCUS 11148 Lambda-PRL2 Arabidopsis thaliana cDNA clone 149B19n7, mRNA
DEFINITION sequence.
ACCESSION T76370
NID 9935370
VERSION T76370.1 GI:935370
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 470)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,
Retzel,E., and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
COMMENT On Apr 14, 1993 this sequence version replaced gi:693132.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI

```

```

Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: T7 dye primer.
FEATURES
    Location/Qualifiers
        1..470
            /organism="Arabidopsis thaliana"
            /strain="var columbia"
            /db_xref="taxon:3702"
            /map="17 cen-qter"
            /clone="149B19n7"
            /note="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not;
            Lambda PRL2 is a cDNA library derived from equal
            quantities of 4 pools of mRNA. The mRNA sources were 1) 7
            day germinated etiolated seedlings; 2) tissue culture
            grown roots; 3) staged plants half with 24 hour light
            cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
            same plants as 3 but aerial tissue (stems, flowers and
            siliques. The vector is BRL's lambda Zip-Lox. The cDNA
            inserts were directionally cloned with Sal-Not arms using
            oligo dT primed cDNA."
BASE COUNT 116 a 104 c 110 g 122 t 18 others
ORIGIN
Query Match      6.9%; Score 132; DB 23; Length 470;
Best Local Similarity 67.3%; Pred. No. 2.3e-13;
Matches 183; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 849 cgcacaactctgtccccaggccaggtaccacaactacgagcgcgccgcttcaa 908
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7 CGCAACATCTTGCTCCACGCTCTGGCTTCCTCATTAGCGCTCGTNCNCTTACAG 66
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 909 caggctgggggtccggcatttcgacctcatcccgacaagggtggagatcgacatcga 968
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67 TGTCTCGAGGTTGCGAAGTTTATCTTCTCCGAGAAGAATGGGAGATTGATCTTNA 126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 969 ctcgctggaatccatcgccgacaagaacaccaccgcatggtcatataaaccccaacaa 1028
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 AGTATCGAAGCATTCGACAGAGAGAAACTGTGGCTGTGTTGTAATTAACCCCAACAA 186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1029 cccgtgcggcagcgtttactctctacgacctctgtccaagggtcgcgaggtggcgaag 1088
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 TCCTCTGGAAATGTCTACTCTCAGCAACCTCTCAAAAAGGTTCAGAGACGCTAGGAA 246
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1089 gtcggaatattgggtgattgctgcagcaggtat 1120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 247 GCTCGGATAATGGTGAATCTCAGACGAAGTAT 278
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
T13684 352 bp mRNA EST 07-JAN-1998
LOCUS 1849 Lambda-PRL2 Arabidopsis thaliana cDNA clone 35D4T7, mRNA
DEFINITION sequence.
ACCESSION T13684
NID 9930780
VERSION T13684.1 GI:930780
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 352)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,
Retzel,E., and Somerville,C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
COMMENT On Apr 14, 1993 this sequence version replaced gi:693132.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI

```

MEDLINE 95148729
COMMENT On Nov 29, 1993 this sequence version replaced gi:430053.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: 17.

FEATURES

Location/Qualifiers

1..352
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/map="4"
/clone="3504T7"
/clone.lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."

BASE COUNT 77 a 61 c 83 g 118 t 13 others
ORIGIN

Query Match 6.3%; Score 121; DB 23; Length 352;

Best Local Similarity 63.9%; Pred. No. 1.6e-11;

Matches 175; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 974 tgaatccatcgccagacaagaccaccccgcatgtcatcataaaccacacccgt 1033

Db 2 TCAGAGCTCTTCGACAGCAAAACACCGTTCCTTGGTTGTATTAACCCNGGTAATCCCTT 61

QY 1034 gcgcagcgtttactctacgaccatctgtccaaagtcgcggaggtggcgaaagctcg 1093

Db 62 GCGGAATNTCTATAGTACCACGATTTGATGAAGATTNCGGAATCGGGGAAAAAAGTAG 121

QY 1094 gaatatggtgtagtgcagaggtatagcggaagctggtctctggcagcgcccgcttca 1153

Db 122 GGTTCCTTGTGATNCTGATGAGGTTTACGGTCATCTTGGTTTGTAGCAAAACCGTTTG 181

QY 1154 tcccaatggaggttttgggcacatcacccctgtgtctcctaggtctctgtccaagt 1213

Db 182 TNCAATNGTGTGTTGGTATGATGTATGTGNCCTGTGCTTACTCTTGGTCTTTATCAAGA 241

QY 1214 catggatgctgctgagtcggcgtctggatgggt 1247

Db 242 GATGGATAGTTCAGGTTTCGACTCGGGTTNGT 275

RESULT 9

D23043

LOCUS 343 bp mRNA EST 20-JUL-1998

DEFINITION RICE2086A Rice callus Oryza sativa cDNA clone R, mRNA sequence.

ACCESSION D23043

NID 9426967

VERSION D23043.1 GI:426967

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; Oryza.

REFERENCE, 1 (bases 1 to 343)

AUTHORS
TITLE
JOURNAL
COMMENT

Sasaki, T. and Minobe, Y.
Rice cDNA from callus
Unpublished (1994)

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".

FEATURES

Location/Qualifiers

1..343
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="R"
/clone.lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."

BASE COUNT 90 a 58 c 80 g 94 t 21 others
ORIGIN

Query Match 6.2%; Score 118.4; DB 20; Length 343;

Best Local Similarity 70.0%; Pred. No. 4.3e-11;

Matches 205; Conservative 0; Mismatches 84; Indels 4; Gaps 4;

QY 1063 tccaaggtcggaggtggtcgaaaggctcggaattgttgatgtgacagaggtat 1122

Db 2 TCCAAGGTGGCAGAGTAGCAAGGAAGCTTGGGATATGGTAATCTACTGATGAGGTGAT 61

QY 1123 ggcaagctggttctgtggcagcgcccttcatcccaatggggaggtgtttgggcacatcac 1182

Db 62 GTANTTTGGTTTTGGGAGTTCNCANTTGTCCCAATNGGTTACTTTGGGCACATCGTA 121

QY 1183 cctgtgctgcattggtgtctgtccaaagtcattgtagtagtgccttggatggcggtctg- 1241

Db 122 CCATAANNANCCNTAGATCGCTATCAAGAGGTGGTAGTCCCGTAGATGCGGACTTGA 181

QY 1241 gatggtagcgtgtacacccccagaaagatcttacagaaactaatctc-tacatca 1299

Db 182 GNTGNTAGCANTAGTGACCCCAAGACACTCTACAGAAACCAAGATTGCANCAATTA 241

QY 1300 attcgaattacctcaatgtctcgacag-accacgaacaccttcattcaggcg 1351

Db 242 ATTACTAATTCCTTAATGGTTCAACTGNATCCAGCANCTTTCATTTCAGGGAG 294

RESULT 10

AI621525/c

LOCUS 283 bp mRNA EST 21-APR-1999

DEFINITION 486092E01.x1 486 - leaf primordia cDNA library from Hake lab Zea

ACCESSION AI621525

NID 94630651

VERSION AI621525.1 GI:4630651

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Poaceae; Zea.

REFERENCE 1 (bases 1 to 283)

AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

Unpublished (1999)

RESULT	12
AA511748	
LOCUS	512 bp mRNA EST
DEFINITION	y29f11.r1 strataegen mouse diaphragm (#937303) Mus musculus cDNA clone IMAGE:1930477 5', similar to gb:X52520_cds1 TYROSINE AMINOTRANSFERASE (HUMAN)), mRNA sequence.
ACCESSION	AA511748
NID	G2249602
VERSION	AA511748.1 GI:2249602
KEYWORDS	EST.
SOURCE	mouse muscle.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Marra,M., Hillier,L., Allien,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theissen,B., Wylier,I., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1398101.

Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:535397
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 366.

```

FEATURES
source
Location/Qualifiers
1. .512
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IWAGE:930473"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/lab_host="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dT. Average
insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATTCGCACGAG 3' ~3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 118 a 122 c 139 g 133 t
ORIGIN
Query Match 4.8%; Score 92.4; DB 34; Length 512;
Best Local Similarity 52.5%; Pred. No. le-06;
Matches 249; Conservative 0; Mismatches 221; Indels 4; Gaps 3

```

Qy	772	gccgacgactcttctccaccgcgcgcgggaccaccaggcgatcgaggttcataatcccggtg	831
Db	4	GCTAAGGATGTCAATTCTGACGAGTGGCTGCGAGTCAGGCATTGAGCTGTGTAGCCGTG	63
Qy	832	ctggccccagacgcgcgcgcgaacatttgtctccccagcgcgggtaccaccaactcaag	891
Db	64	TTGCCC---AATCTGGACAGAATCTCATTCGGAGGCCGGGTTTCCCTCTACAGG	120
Qy	892	gcgcgcgcgcgcgtttcaacagcgctggagctcgcggcatcttcgacctcatccccgcacaagggg	951
Db	121	ACATTGGCTGAGTCTATGGGATTGAGTCAAGCTCTACAATCTATTGCCTGAGAAGTCT	180
Qy	952	tgggagatcgacatcgactcgtcggaatcccatcgccgaagaacaccacgcgcgatggtc	1011

Db	181	TGGGAATTGATCTAAACAACACTGGAATCTCTGATCGACGAAAAACACAGCTTGTTCTCGTG	240		
Qy	1012	atcataaaccccccaaacccctgcgcagcgtttactcctaagaccattctgtccaagttc	1071		
Db	241	GTCACAACCCGTCGCAATCCCTGTGGCTCTGTGTTTCAGTAAGCGGCACCTTCAGAGAAT	300		
Qy	1072	gcgaggtggcgaaaggctcggaatatgtgtgattgtcgcagaggtatcacgcaagctg	1131		
Db	301	TGGCAGTGGCTGAAAGGCAATGGCTCCCATCTTAGCCGATGAGATCTATGGTGACATG	360		
Qy	1132	gttctgggcagcgcccgcttcattcccaatgggagtgtttgggcacatcacccctgtgctg	1191		
Db	361	GTGTTTCAGATTGCAMATATGAACCAATGCCACCCCTCAGCACCAATGTCCCCATCCCTG	420		
Qy	1192	tcctagggctctgtccaagtcatcgtagtcgtcgatgcggcctggatgg	1245		
Db	421	TCCTGTGTGGGGCTG-GCAAGCGTGGCTGTTTCTGGGCTGAGGTGGGCTGG	473		
RESULT 13					
Z38047/c					
LOCUS	Z38047	405 bp	mrna EST 12-OCT-1994		
DEFINITION	AT54207 Versailles-VB Arabidopsis thaliana cDNA clone VBWMA10 3',				
RNA sequence.					
ACCESSION	Z38047				
NID	g557616				
VERSION	Z38047.1	GI:557616			
KEYWORDS	EST.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
Eudicotyledons; Magnoliophyta; Magnoliopsida; Eudicotyledons; core					
eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae;					
Arabidopsis.					
REFERENCE	1	(bases 1 to 405)			
AUTHORS	CNRS.				
TITLE	The Arabidopsis thaliana transcribed genome: the GDR cDNA program				
JOURNAL	Unpublished (1996)				
COMMENT	Contact: Desprez T., Anselem J., Chiapello H., Rouze P., Caboche M., Hofte H.				
	INRA Versailles				
	Laboratoire de Biologie Cellulaire				
	Route de Saint-Cyr, 78026 Versailles Cedex, France				
	Email: thieriev@versailles.inra.fr.				

FEATURES	source	Location/Qualifiers
BASE COUNT	132 a	80 c 73 g 117 t 3 others
ORIGIN		
	1. .405	
	/organism="Arabidopsis thaliana"	
	/strain="ecotype Columbia"	
	/db_xref="taxon:3702"	
	/clone_lib="VBVMA10"	
	/clone_lib="Versailles-VB"	
	/note="vector: pBluescript; tissue=whole seedlings of A. thaliana ecotype Columbia; clone_library=Versailles-VB; Cloning vector: pBluescript (Stratagene); Physiological conditions: in vitro-grown etiolated seedlings, 5 days old."	

[illegible]

